

AA003705 331 bp mRNA  
mg1a02.r1 Soares mouse embryo NBxEL3.5 14.5 Mus musculus cDNA  
clone IMAGE:437450 5', mRNA sequence.  
AA003705  
AA003705.1 GI:1447205  
EST.  
house mouse.  
Mus musculus  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,  
Geissel, S., Kucaba, T., Lacy, M., Le M., Martin, J., Morris, M.,  
Schallenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,  
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R., and  
Waterston, P.

REFERENCE  
AUTHORS

[illegible]

Contact: Maria M/Mouse EST Project  
WashU-HHMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810

Email: mouseest@watson.wustl.edu  
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
MGI:262786  
Seq primer: ETPRimer

SEQ ID NO: 55

Tue Apr 30 14:18:38 2002

us-09-248-]

FEATURES  
SOURCE

```

Location/Qualifiers
1. .331
  /organism="Mus musculus"
  /strain="C57BL/6J"
  /db_xref="taxon:10090"
  /clone_image="437450"
  /clone_lib="Soares mouse embryo NbME13.5 14.5"
  /sex="unknown"
  /issue_type="embryo"
  /dev_stage="13.5-14.5dpc total fetus"
  /lab_host="DH10B"
  /note="Vector (Pharmacia) with a modified
polylinker. Site.1: Not I; Site.2: Eco RI;
Site.3: Not I - oligo(dT) primer [5'
TCTTACCAATCTGAAGTGGCGCGCGGAATTTTTTTTTT
T 3'], on equal amounts of mRNA from 2
13.5dpc and 2
14.5dpc embryos [total RNA provided by
Munro K, Wayne
State Univ., from 2]; double-stranded cDNA was ligated to
Eco RI adaptors (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of the modified
p773 vector. Library went through one round of
normalization, and was constructed by Bento Soares and
M. Fatima Bonaldo."

```

BASE COUNT	72 a	65 c	73 g	121 t
ORIGIN				

Query Match	2.5%;	Score 23;	DB 10;	Length 331;
Best Local Similarity	100.0%;	Pred. No. 1.3e+03;		
Conservative	0;	Mismatches 0;	Indels 0;	Gaps 0;

911 tatctqccaaaaa 933

298 TATCTGCCAAAAA 320



XX 31-OCT-1996; 96US-0029960.  
 PR (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PA Barash SC, Choi GH, Dillon FJ, Dougherty BA, Fannon M;  
 PI Kunsch CA, Rosen CA;  
 XX WPI; 1998-27225/24.  
 DR  
 XX Computer-readable medium with recorded Streptococcus pneumoniae  
 PT polynucleotide sequences - useful in diagnostic kits and assays, and  
 PT pharmaceutical compositions and vaccines for Streptococcus  
 PT pneumoniae.  
 XX  
 PS Claim 1; Page 608-617; 1409pp; English.  
 XX  
 CC The present invention describes a computer readable medium which has  
 CC the nucleotide sequences SEQ ID NO:1 to 391 (AAV52134 to AAV52524)  
 CC recorded on it, or a representative fragment or a sequence at least 95%  
 CC identical to SEQ ID NO: 1 to 391. The nucleotide sequences depicted in  
 CC SEQ ID NO:1 to 391 (AAV52134 to AAV52524) are genomic fragments from  
 CC Streptococcus pneumoniae. The present invention also describes an  
 CC isolated nucleic acid molecule encoding a homologue of any of the  
 CC fragments of the S.pneumoniae genome (SEQ ID NO:1 to 391) where the  
 CC nucleic acid molecule is produced by a process comprising: (a) screening  
 CC a genomic DNA library using as a probe a target sequence defined by any  
 CC of the sequences in SEQ ID NO:1 to 391, identifying members of the  
 CC library which contain sequences that hybridize to the target sequence and  
 CC isolating the nucleic acid molecules from the members; or (b) isolating  
 CC mRNA, DNA or cDNA produced from an organism, amplifying nucleic acid  
 CC molecules whose nucleotide sequence is homologous to amplification  
 CC primers derived from the fragment of the S. pneumoniae genome to prime  
 CC the amplification and isolating the amplified sequences. The computer  
 CC readable medium can be used in a computer-based system for identifying  
 CC fragments of the S. pneumoniae genome of commercial importance, or  
 CC expression modulating fragments of the S. pneumoniae genome. Products  
 CC from the present invention can be used in diagnosis kits and assays, and  
 CC pharmaceutical compositions and vaccines for S. pneumoniae.  
 XX  
 SQ Sequence 16535 BP; 5001 A; 3914 C; 3102 G; 4518 T; 0 other;

Query Match 2.8%; Score 17; DB 19; Length 16535;  
 Best Local Similarity 100.0%; Pred. No. 86;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 474 ttcaaggatgcccaagaa 490  
 |||||  
 Db 451 ttcaaggatgcccaagaa 467

SEQ.ID.NO:60

RESULT 43  
 AAV52207  
 ID AAV52207 standard; DNA; 16535 BP.  
 XX  
 AC AAV52207;  
 XX  
 DT 23-OCT-1998 (first entry)  
 XX  
 DE Streptococcus pneumoniae genome fragment SEQ ID NO:74.  
 XX  
 KW Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;  
 KW computer readable medium; vaccine; pharmaceutical composition; ds.  
 XX  
 OS Streptococcus pneumoniae.  
 XX  
 PN WO9818931-A2.  
 XX  
 PD 07-MAY-1998.  
 XX  
 PF 30-OCT-1997; 97WO-US19588.

Tue Apr 30 14:18:55 2002

US-08-998-416-136

Query Match 5.0%; Score 68; DB 4; Length 703;  
 Best Local Similarity 100.0%; Pred. No. 8.7e-18;  
 Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1299 gccacgcggtgagctccagcttttgcctttagtgagggtaattgcgcgttgcc 1358  
 |||||  
 DB 158 GCCACGCGGTGGAGCTCCAGCTTTTGTCCCTTTAGTGAGGGTTAATTGCGCGCTTGC 217

QY 1359 gtaatacat 1366  
 |||||  
 DB 218 GTAATCAT 225

SEQ ID NO: 61

RESULT 1  
 US-08-998-416-136  
 ; Sequence 136, Application US/08998416  
 ; Patent No. 6239264  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Philippsen, Peter  
 ; APPLICANT: Pohlmann, Rainer  
 ; APPLICANT: Steiner, Sabine  
 ; APPLICANT: Mohr, Christine  
 ; APPLICANT: Wendland, Jurgin  
 ; APPLICANT: Knechtle, Philipp  
 ; APPLICANT: Rebischung, Corinne  
 ; TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSTYPII  
 ; TITLE OF INVENTION: AND USES THEREOF  
 ; NUMBER OF SEQUENCES: 1152  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: No. 6239264artis Corporation  
 ; STREET: 3054 Cornwallis Road  
 ; CITY: Research Triangle Park  
 ; STATE: No. 6239264th Carolina  
 ; COUNTRY: USA  
 ; ZIP: 27709  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/998,416  
 ; FILING DATE: 24-DEC-1997  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: CH 0016/97  
 ; FILING DATE: 31-DEC-1996  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Meigs, J. Timothy  
 ; REGISTRATION NUMBER: 38,241  
 ; REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1975  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 919-541-8587  
 ; TELEFAX: 919-541-8689  
 ; INFORMATION FOR SEQ ID NO: 136:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 703 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: DNA (genomic)  
 ; ORIGINAL SOURCE:  
 ; ORGANISM: PAG1051UP



```

RESULT 30
US-08-578-649-4
; Sequence 4, Application US/08578649
; Patent No. 5770366
; GENERAL INFORMATION:
; APPLICANT: Ulrich Bogdan
; APPLICANT: Reinhard Buttner
; APPLICANT: Brigitte Kaluza
; TITLE OF INVENTION: MELANOMA-INHIBITING PROTEIN
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch 1.44 Mb storage diskette
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/578,649
; FILING DATE: 29-July-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE P 43 24 247.2
; FILING DATE: 20-July-1993
; ATTORNEY/AGENT INFORMATION:

```

```

NAME: Andrew L. Tiajolloff
REGISTRATION NUMBER: 31,575
REFERENCE/DOCKET NUMBER: BOER 1035-PFF/ALT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 688-9200
TELEFAX: (212) 838-3884
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 581 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cdna
FEATURE:
NAME/KEY: CDS
LOCATION: 110..499
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 110..178
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 179..499
US-08-578-649-4

```

Query Match 1.9%; Score 18; DB 1; Length 581;  
Best Local Similarity 100.0%; Pred. No. 35;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 907 ggcaaaaaaaaaaaaaaaaaa 924  
 |||||  
 Db 563 GGCAAAAAAAAAAAAAAAAA 580

SEQ ID NO: 63

RESULT 17  
AAV55038  
ID AAV55038 standard; cDNA; 5232 BP.  
XX  
AC AAV55038;  
XX  
DT 13-NOV-1998 (first entry)  
XX  
DE Human XIAP coding sequence.  
XX  
KW Inhibitor of apoptosis protein; apoptosis enhancer; NAIP polypeptide;  
KW proliferative disease; IAP; therapy; cancer; human; XIAP protein; ss.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT CDS 34..1527  
FT /\*tag= a  
FT /product= XIAP  
XX  
PN WO9835693-A2.  
XX  
PD 20-AUG-1998.  
XX  
PF 13-FEB-1998; 98WO-IB00781.  
XX  
PR 13-FEB-1997; 97US-0800929.  
XX  
PA (UYOT-) UNIV OTTAWA.  
XX  
PI Baird S, Korneluk R, Liston P, Mackenzie AE, Pratt C;  
PI Tsang B;  
XX  
DR WPI; 1998-467164/40.  
DR P-PSDB; AAW69294.  
XX  
PT Inducing apoptosis in proliferative mammalian cells with inhibitor  
PT of IAP or NAIP polypeptide - also methods for prognosis based on  
PT presence of IAP and NAIP, specifically applied to cancers involving  
PT p53 mutations  
XX  
PS Claim 13; Fig 1; 147pp; English.  
XX  
CC This sequence encodes the human XIAP protein, which is a inhibitor of  
CC apoptosis protein (IAP), and can be used in the method of the invention.  
CC The method is for enhancing apoptosis in cells from a mammal with  
CC proliferative disease by treatment with a compound that inhibits  
CC biological activity of an IAP or NAIP polypeptide. The inhibitory  
CC compounds are used to treat proliferative diseases, specially cancers of  
CC ovary, breast, pancreas, lymph nodes, skin, blood, lung, brain, kidney,  
CC liver nasopharynx, thyroid, central nervous system, prostate, colon,  
CC rectum, cervix or endometrium, particularly to increase their sensitivity  
CC to chemotherapeutic agents. High levels of the IAP or NAIP proteins are  
CC detected in many cancers and are associated with poor prognosis,  
CC resistance to chemotherapeutic agents and mutations in p53 (it is  
CC suggested that wild-type p53 suppresses transcription of the IAP or NAIP  
CC genes). Transgenic animals are used for testing the effects of antisense  
CC oligonucleotides and for screening for the inhibitors.  
XX  
SQ Sequence 5232 BP; 1579 A; 861 C; 1062 G; 1728 T; 2 other;

Query Match 1.9%; Score 21; DB 19; Length 5232;  
Best Local Similarity 100.0%; Pred. No. 1.6;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 459 agtttttctaatccaagaagg 479  
|||||  
Db 2498 agtttttctaatccaagaagg 2518

RESULT 18

SEQ ID NO 64

RESULT 7  
AAQ59633/C  
ID AAQ59633 standard; cDNA; 319 BP.  
XX  
AC AAQ59633;  
XX  
DT 16-MAR-1994 (first entry)  
XX  
DE Human brain Expressed Sequence Tag EST01493.  
XX  
KW Gene transcription product; genetic markers; tagging; in vivo;  
KW transcription; mapping; locations; chromosomes; chromosomal; ss.  
XX  
OS Homo sapiens.  
XX  
PN W09316178-A.  
XX  
PD 19-AUG-1993.  
XX  
PF 12-FEB-1993; 93WO-US01294.  
XX  
PR 12-FEB-1992; 92US-0837195.  
XX  
PA (USSH ) US DEPT HEALTH & HUMAN SERVICE.  
XX  
PI Adams MD, Moreno RF, Venter CJ;  
XX  
DR WPI; 1993-272882/34.  
XX  
PT Enriched oligonucleotides and corresp. sequences - used as  
PT markers for human genes transcribed in-vivo, facilitate tagging  
XX of most human genes  
PS Example 4; Page 207; 500pp; English.  
XX  
CC The Expressed Sequence Tag was isolated from a human brain cDNA  
CC library as part of a large set of ESTs which can be used as markers  
CC for human genes transcribed in vivo. They can be used to facilitate  
CC tagging of most human genes, for mapping locations of expressed genes  
CC on chromosomes, for individual or forensic identification, for mapping  
CC locations of disease-associated genes, for identification of tissue  
CC type, and for prepn. of antisense sequences, probes and constructs.  
CC EST01493 has a "poor" coding probability as evaluated using the  
CC coding-region prediction program CRM. See also AAQ59041-Q61440.  
XX  
SQ Sequence 319 BP; 118 A; 42 C; 75 G; 82 T; 2 other;  
  
Query Match 2.7%; Score 27; DB 14; Length 319;  
Best Local Similarity 100.0%; Pred. No. 0.12;  
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 341 cttattttcacttaacataatgacctcc 367  
Db 90 CTTATTCTACTTAACATAATGACCTCC 64  
  
RESULT 8

SEQ ID NO 65

RESULT 1  
US-09-040-984-27  
; Sequence 27, Application US/09040984  
; Patent No. 6210883  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Wang, TongTong  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS  
; TITLE OF INVENTION: OF LUNG CANCER  
; NUMBER OF SEQUENCES: 86  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SEED and BERRY LLP  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: WA  
; COUNTRY: USA  
; ZIP: 98104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSEQ for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/040,984  
; FILING DATE: 18-MAR-1998  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Maki, David J.  
; REGISTRATION NUMBER: 31,392  
; REFERENCE/DOCKET NUMBER: 210121.456  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 206-622-4900  
; TELEFAX: 206-282-6031  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 27:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 654 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-09-040-984-27

Query Match 55.5%; Score 319; DB 4; Length 654;  
Best Local Similarity 100.0%; Pred. No. 2.7e-129;  
Matches 319; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 acttgatataaaaaggatatccataatgaatattttatactgcatcctttacattagcca 60  
| | | | |  
Db 88 ACTTGATATAAAAAGGATATCCATAATGAATATTTTATACTGCATCCTTTACATTAGCCA 147

Qy 61 ctaaatacgttattgcttgatgaagacctttcacagaatcctatggattgcagcatttca 120  
| | | | |  
Db 148 CTAATAACGTTATTGCTTGATGAAGACCTTTCACAGAATCCTATGGATTGCAGCATTTC 207  
| | | | |  
Qy 121 ctggctacttcatacccatgccttaagaggggagctttctcaaaagcagaacatgcc 180  
| | | | |  
Db 208 CTTGGCTACTTCATACCCATGCCTTAAGAGGGGCGAGTTTCTCAAAAGCAGAAACATGCC 267  
| | | | |  
Qy 181 gccagttctcaagttttctcctcctaactccatttgaatgtaagggcagctggccccaatg 240  
| | | | |  
Db 268 GCCAGTTCTCAAGTTTCTCCTCAACTCCATTTGAATGTAAGGGCAGCTGCCCCCAATG 327  
| | | | |  
Qy 241 tggggagggtccgaacattttctgaattcccatcttctgttcgaggctaaatgacagttt 300  
| | | | |  
Db 328 TGGGGAGGTCCGAACATTTCTGAATCCCATTTCTGTTCGCGGCTAAATGACAGTTT 387  
| | | | |  
Qy 301 ctgtcattacttagattcc 319  
| | | | |  
Db 388 CTGTCATTACTTAGATTCC 406

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 29, 2002, 19:34:49 ; Search time 8020.64 Seconds  
(without alignments)  
1570.670 Million cell updates/sec

Title: US-09-248-178-60  
Perfect score: 602  
Sequence: 1 tgaagacgcgcgggtggag.....tcacgcggaggttcagac 602

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues  
Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

- GenEmbl:\*
- 1: gb\_ba.\*
  - 2: gb\_htg.\*
  - 3: gb\_in.\*
  - 4: gb\_om.\*
  - 5: gb\_ov.\*
  - 6: gb\_pat.\*
  - 7: gb\_ph.\*
  - 8: gb\_pl.\*
  - 9: gb\_pr.\*
  - 10: gb\_ro.\*
  - 11: gb\_sts.\*
  - 12: gb\_sy.\*
  - 13: gb\_un.\*
  - 14: gb\_vi.\*
  - 15: em\_ba.\*
  - 16: em\_fun.\*
  - 17: em\_hum.\*
  - 18: em\_in.\*
  - 19: em\_mu.\*
  - 20: em\_om.\*
  - 21: em\_or.\*
  - 22: em\_ov.\*
  - 23: em\_pat.\*
  - 24: em\_ph.\*
  - 25: em\_pl.\*
  - 26: em\_ro.\*
  - 27: em\_sts.\*
  - 28: em\_un.\*
  - 29: em\_vi.\*
  - 30: em\_htg\_hum.\*
  - 31: em\_htg\_inv.\*
  - 32: em\_htg\_other.\*
  - 33: em\_htgo\_inv.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
------------	-------	-------	--------	----	-------------

NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [mgc@nci.nih.gov](mailto:mgc@nci.nih.gov)  
Tissue Procurement: ATCC

ALIGNMENTS

RESULT 1  
BC011675

LOCUS  
DEFINITION

ACCESSION  
VERSION  
KEYWORDS

SOURCE  
ORGANISM

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

REMARK  
COMMENT

BC011675 2168 bp mRNA linear PRI 02-AUG-2001  
Homo sapiens, similar to JMI protein, clone MGC:15381  
IMAGE:4299954, mRNA, complete cds.

BC011675  
BC011675.1 GI:15079716  
MGC.  
human.

Homo Sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 2168)  
Strausberg, R.

Direct Submission  
Submitted (30-JUL-2001) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [mgc@nci.nih.gov](mailto:mgc@nci.nih.gov)  
Tissue Procurement: ATCC

BC011675 Homo sapi  
AJ005890 Homo sapi  
BC000972 Homo sapi  
AF229637 Mus muscu  
BC011195 Mus muscu  
AF235097 Homo sapi  
AF277994 Mus muscu  
AC097025 Rattus no  
AF277991 Mus muscu  
AF277992 Mus muscu  
AY058359 Drosophil  
AC014936 Drosophil  
AC010751 Drosophil  
AC006933 Drosophil  
AE003526 Drosophil  
AC016204 Homo sapi  
AC011767 Homo sapi  
AC105301 Mus muscu  
AC105305 Mus muscu  
Continuation (17 o  
AC093352 Mus muscu  
AC091951 Homo sapi  
AC016768 Homo sapi  
AC099414 Mus muscu  
AC090656 Mus muscu  
AC105317 Homo sapi  
AF429315 Homo sapi  
AC067854 Homo sapi  
AC104807 Homo sapi  
S76368 ORF 5' of E  
M86409 Herpesvirus  
X64346 Herpesvirus  
I66494 Sequence 14  
AC002304 Genomic s  
AC002328 Genomic s  
AC018348 Homo sapi  
AC091304 Homo sapi  
AC023310 Homo sapi  
AJ289860 Lampetra  
AF192756 Kaposi's  
AC095672 Rattus no  
X60297 A.californi  
U72849 Homo sapien  
U53786 Homo sapien  
AC040980 Homo sapi

571 94.9 2168 9 BC011675  
571 94.9 2228 9 HSA005890  
571 94.9 2280 9 BC000972  
459.6 76.3 2179 10 AF229637  
459.6 76.3 2194 10 BC011195  
121.8 20.2 130788 9 AF235097  
101.2 15.8 30858 10 AF277994  
88.2 11.7 162389 2 AC097025  
84.2 14.0 3681 10 AF277991  
84.2 14.0 3739 10 AF277992  
77.5 12.9 1828 3 AY058359  
77.6 12.9 97259 2 AC014936  
77.6 12.9 181437 3 AC010751  
77.5 12.9 186002 3 AC006933  
77.6 12.9 280887 3 AE003526  
57.2 9.5 142277 2 AC016204  
56.4 9.4 215046 2 AC011767  
56.2 9.3 206976 2 AC105301  
56.2 9.3 221657 2 AC105305  
48.4 8.0 110000 2 LMFLCHR34\_16  
48.2 8.0 229329 2 AC093352  
46.5 7.7 159971 2 AC091951  
46 7.6 184451 9 AC016768  
46 7.5 219521 2 AC099414  
45.8 7.6 211973 2 AC090656  
45.6 7.6 173294 2 AC105317  
45 7.5 125020 9 AF429315  
45 7.5 203984 2 AC067854  
44.4 7.4 183552 2 AC104807  
44 7.3 3720 14 S76368  
44 7.3 43658 14 HSV3PRGEN  
44 7.3 112930 14 HSGEND  
43.8 7.3 7218 6 I66494  
43.5 7.2 106320 8 AC002304  
43.5 7.2 109171 8 AC002328  
43.4 7.2 178022 2 AC018348  
43.4 7.2 179337 2 AC091304  
43.2 7.2 200019 2 AC023310  
43 7.1 2012 5 LFL289860  
43 7.1 2931 14 AF192756  
42.8 7.1 229706 2 AC095672  
42.4 7.0 3353 3 ACIFPA  
42.4 7.0 4513 9 HSAPEVPL7  
42.4 7.0 6457 9 HSD53786  
42.4 7.0 150244 2 AC040980

cDNA Library Preparation: Rubin Laboratory  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILLNL)  
 DNA Sequencing by: National Institutes of Health Intramural  
 Sequencing Center (NISC),  
 Gaithersburg, Maryland;  
 Web site: <http://www.nisc.nih.gov/>  
 Contact: [nisc.mgc@hqr1.nih.gov](mailto:nisc.mgc@hqr1.nih.gov)  
 Shcherchenko, Y., Wetherby, K.D., Beckstrom-Sternberg, S.M.,  
 Benjamin, B., Blakesley, P.W., Bouffard, G.G., Brinkley, C., Brooks, S.,  
 Dietrich, N.L., Guan, X., Gupta, J., Ho, S.-L., Karlins, E., Ledas, P.R.,  
 Lim, M., Maduro, O.L., Masiello, C., Mastrian, S.D., McCloskey, J.C.,  
 McDowell, J., Pearson, R., Snyder, B., Stantripop, S., Thomas, P.J.,  
 Tionsong, E.E., Touchman, J.W., Tsugeon, C., Vogt, J.L., Walker, M.A.,  
 Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found  
 through the I.M.A.G.E. Consortium/ILLNL at: <http://image.llnl.gov>  
 Series: IPAL plate: 26 Row: n Column: 14  
 This clone was selected for full length sequencing because it  
 passed the following selection criteria: Similarity but not  
 identity to protein.

## FEATURES

source  
 1..2168  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="MGC:15381 IMAGE:4299954"  
 /tissue\_type="Pancreas, epithelioid carcinoma"  
 /clone\_lib="NIH\_MGC\_42"  
 /lab\_host="DH10B-R"  
 /note="Vector: pOTB7"  
 11..1894  
 /codon\_start=1  
 /product="Similar to JM1 protein"  
 /protein\_id="AAH11675.1"  
 /db\_xref="GI:15079717"  
 /translation="MEEADRLIHSLRQAGTAVPDVOTLRAFTTELVEAVVRLV  
 IPVAGSLPLPLAMSLAQACMDLGYPLEGFONFLYSEDLRLLL  
 FLAEHLPTDASEDADPADGSAILLRATGQIRDLALWPPPHLRTKLOHLOQSAL  
 QKPFHLSRLVPELSRGEFQASPLLPVTPQVPGVRSVLSLLEHHAOLCCOIT  
 GRDPCDEWHRTSRLPQEDTRAQORLQKLTLEHLRQSLGLGAPIQAPDLGELL  
 QAWGAKGTWPKGRFTSEKFTFLEPQAQATQVSDVPATSRPEQVITAAQOELL  
 ESLRQELGNSIEVEADMKTLGVSFVQAESECRHSLSTAEREOLPIKSPAVEL  
 LDGTANLAKLVVENSQORVILHAGQWEKRVPLLAFLYRHLKLODCLPFLSSRPL  
 AEIOELHOSVRAAEARKKEVYKOLMSELTLPDVSRLAYTORILEIVINIKQK  
 EEITLTSLDKELQELNSLGLKDRITFAVDLKYDKAKDDAVFKYKILALHEN  
 CSQLOITIEDTGTIMREVRDLPEQIETELGKKILSNLEKIREYRALRQETAGILGRV  
 REA"  
 BASE COUNT 441 a 563 c 675 g 389 t  
 ORIGIN

Query Match 94.9%; Score 571; DB 9; Length 2168;  
 Best Local Similarity 98.7%; Pred. No. 2.le-121;  
 Matches 597; Conservative 0; Mismatches 5; Indels 3; Gaps 2;  
 QY 1 tgaagaccgcgcggtgagactgtcccgatggactgccaaccitgccaaagctcacc 60  
 |||||  
 DB 1164 TGAAGACCCGCGGTGGAGACTGCTGCCGATGGGACTGCCACCTTGCCAACTGCAGC 1223  
 |||||  
 QY 61 ttgtgtgagaatagtgccacagcggtgcatccactggcgggtcagtgaggagaagcacc 120  
 |||||  
 DB 1224 TTGTGTGGAGAAATAGTGCACAGCGGTGATCCACITTCGCGGGTCACTGGGAGAGCACC 1283  
 |||||  
 QY 121 ggtgccca--tctctggaataccgcca-ctccgaagctgcaggaattgcagagaactgg 177  
 |||||  
 DB 1294 GGGTCCCACTCTCTGCTGAGTACCGCCACCTCCGAAAGCTGTCAGAGATTCGAGAGAGCTGG 1213  
 |||||  
 QY 178 aatctctcgaagcgtggcagagatccaaagaactgcaccagagatgtccggcgcggtcgtg 217  
 |||||  
 DB 1344 AATCTCTCGACGCTGCGCAGAGATCCAAAGACTGCACACAGAGTGTCCGGCGGCTGCTG 1103  
 |||||  
 QY 238 aagagggccgcagagaagagaggtctataagcagctgatctcagaactcagaactctac 297  
 |||||

DB 1404 AAGAGGCCCGCAGGAGGAGGAGGTCTATATAGCAGCTGATGTCAGAGCTGGAGACTCTGC 1463  
 QY ccagagatgtgtcccgctgacccacccagcgcatcctctggagatcgtgggcaaatcc 357  
 |||||  
 DB 1464 CCAGAGATGTCTCCGGCTGCGCTACACCCAGCGCATCTCTGGAGATCGTGGGCAACATCC 1523  
 QY ggaagcagaagaagagatcaccaagatcttctgtatcacgaagagagcttcagagaagaa 417  
 |||||  
 DB 1524 GGAAGCAGAAGAGAGATACCAAGATCTTGTCTGTATACCAAGAGCTTCAGAAAGNAA 1583  
 QY tcaactcctatctctggaagctgacccggacgtttggcggtgactgagagcttctgttca 477  
 |||||  
 DB 1584 TCAACTCCTATCTGGGAAGCTGGACCGGACGTTTCGCGGTGACTGATGAGCTTGTGTCA 1643  
 QY aggatgccaaagagagatcttctggaagagcctataagatactgaactctctgcacg 537  
 |||||  
 DB 1644 AGGATGCCAAGAGGACCATGCTGTTCGGAAGGCGCTATAGTATCTAGTCTCTGCACG 1703  
 QY aagaatgcagccagctcattccagaccatcgaggagcacagcaccatcatcgcggaagttc 597  
 |||||  
 DB 1704 AGAACTGCAGCCAGCTCATCCAGACCATCGAGGACACAGGCCCATCATCGGGAGGTTTC 1763  
 QY ggaac 602  
 |||||  
 DB 1764 GAGAC 1768  
 |||||  
 RESULT 2  
 HSA005890 HSA005890 2228 bp mRNA linear PRI 15-MAY-1998  
 LOCUS Homo sapiens mRNA for JM1 protein, complete CDS (clone  
 DEFINITION LLNLc110M011107 (RZPD Berlin)and LLNLc110K214007 (RZPD Berlin)).  
 ACCESSION AJ005890  
 VERSION AJ005890.1 GI:3114815  
 KEYWORDS JM1 gene.  
 SOURCE human  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 2228)  
 AUTHORS Strom, T.M., Nyakatura, G., Hellebrand, H., Drescher, B., Rosenthal, A.  
 and Meindl, A.  
 TITLE Transcription map in Xp11.23  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 2228)  
 AUTHORS Strom, T.M.  
 TITLE Direct Submission  
 JOURNAL Submitted (30-APR-1998) Abteilung Medizinische Genetik, Goethestr.  
 29, Muenchen 80336, Germany  
 FEATURES  
 Location/Qualifiers  
 1..2228  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="X"  
 /map="p11.23"  
 /clone="LLNLc110M011107 (RZPD Berlin)"  
 /clone="LLNLc110K214007 (RZPD Berlin)"  
 /tissue\_type="peripheral blood leukocytes"  
 87..1970  
 /gene="JM1"  
 /gene="JM1"  
 /gene="JM1"  
 /codon\_start=1  
 /protein\_id="CAA06747.1"  
 /db\_xref="GI:3114816"  
 /translation="MEEADRLIHSLRQAGTAVPDVOTLRAFTTELVEAVVRLV  
 IPVAGSLPLPLAMSLAQACMDLGYPLEGFONFLYSEDLRLLL  
 FLAEHLPTDASEDADPADGSAILLRATGQIRDLALWPPPHLRTKLOHLOQSAL  
 QKPFHLSRLVPELSRGEFQASPLLPVTPQVPGVRSVLSLLEHHAOLCCOIT  
 GRDPCDEWHRTSRLPQEDTRAQORLQKLTLEHLRQSLGLGAPITAAQOELL  
 QAWGAKGTWPKGRFTSEKFTFLEPQAQATQVSDVPATSRPEQVITAAQOELL  
 ESLRQELGNSIEVEADMKTLGVSFVQAESECRHSLSTAEREOLPIKSPAVEL  
 LDGTANLAKLVVENSQORVILHAGQWEKRVPLLAFLYRHLKLODCLPFLSSRPL  
 AEIOELHOSVRAAEARKKEVYKOLMSELTLPDVSRLAYTORILEIVINIKQK  
 EEITLTSLDKELQELNSLGLKDRITFAVDLKYDKAKDDAVFKYKILALHEN  
 CSQLOITIEDTGTIMREVRDLPEQIETELGKKILSNLEKIREYRALRQETAGILGRV  
 REA"  
 gene  
 CDS

LPDGTANLAKQLVWVNSNORVTHLAGOWEKHPVPLLAEYRHLPRLODCRELESSRRL  
AEIOELHOSVRAAEAEARKEEYVKNLSELEITL PVDKSLAYTORIEVLVYNBROK  
EETIKHSOTKEKEINSLSKIDRTFAVTQIF VPKDAKIDAVAKAYKYLVAAL HEN  
CSALIOT IEDTIGTIMREVDLEFEO IETELCKKTLI SNLEK IREDYPALROENAGHILGRV  
CSALIOT

BASE COUNT	REA	ORIGIN
438 a	597 c	690 g
		403 t

Query Match 94.9%; Score 571; DB 9; Length 2228;  
Best Local Similarity 98.7%; Pred. NO. 2.le-121;  
Matches 597: Conservative 0; Mismatches 5; Indels

Qy	1	tgaagaaccgcgcgtatgagctgtcccgatggagactgccaaaccttgcacaagctgcaggc	50
Db	1240	TGAAGAGCCGCGGGTGGAGCTGTGCGCGATGGAGACTGCCAACTTTCCTCAAGTGTGAGC	1299
Qy	61	tgtgtgtggaaataatgtcccaacgcggtatccactctggcgagttcaatggaaagaagcacc	120
Db	1300	TTGTGTGGAGAAATAGTGCACAGCGGGTCATCCACTTGGCGGGTCAGTGGAGGAAGACCC	1359
Qy	121	gggtccca--tctctgtgaataccgcca-ctccgaagctgcaggattgcagaagaatgg	177
Db	1360	GGGTCCCACTCTCGCTTGAGTACCGCCACCTCCGAAAGCTCAGAGATTCAGAGAGTGG	1419
Qy	178	aattctctcaacggtgcgcagagatccaagaactgcaccagaagtctccggcgctctgt	237
Db	1420	AATCTTCTCAGCGCTGSCAGAGATCCAGAAGCTGCCACAGAGTTCGGGCGCTCTGTG	1479
Qy	238	agaagggccgcagaaagagaggctctataagcagctgaatctcaagcctgaagactctgc	297
Db	1480	AAGAGGCCGCGAGGAAGGAGGAGGTCTATAGCAGCTGATGTCAGAGCTGGAGACTTGC	1539
Qy	298	ccagagatgtgtcccgctggcctacaccagcgatccttgagaatcgatggcgacaatcc	357
Db	1540	CCAGAGATGTCTCCGCTTGGCCTACACCCAGCGCATCTCTGGAGATCTGTGGGCAACATCC	1599
Qy	358	ggaagcagaagaagagatcacaagatcttgtgatacgaagaagacttcagaagaaaa	417
Db	1600	GGAGCAGAGGAAGAGATCACCAGATCTTGCTGATACCAGAGAGCTTCAGAAAGGAAA	1659
Qy	418	tcaactccctatctggaaagctggaaccggacgtttgcggtgaactgatgagcttgtgttca	477
Db	1660	TCAACTCCCTATCTGGGAAGCTGGACCGGAGCTTGGCGGTGACTGTATGAGTCTGTGTCA	1719
Qy	478	agatgtccagaagagcgatgctgttccgaagcgctataaatacttaagctactctgcacg	537
Db	1720	AGGATGCCAAGAGGACGATGCTGTTTGGGAAGGCTATAGTATCTAGTCTCTGTACG	1779
Qy	538	agaactgcagccagctcatccagaccatcaggaacagagaccatcatcgcgagagttc	597
Db	1780	AGAACTGCAGCCAGCTCATTCAGACCATCGAGGACAGGGCACCATATCGGGAGGTTTC	1839
Qy	598	gagac	602
Db	1840	GAGAC	1844

RESULT	3
BC000972	
LOCUS	2280 bp mRNA linear PFI: 12-JUN-2001
DEFINITION	Homo sapiens, JM1 protein, clone MGC:5137 IMAGE:3449051. mRNA, complete cds.
ACCESSION	BC000972
VERSION	BC000972.2 GI:12903024
KEYWORDS	MGC.
SOURCE	human.
ORGANISM	Homo sapiens
	Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE	1 (bases 1 to 2280)
AUTHORS	Strausberg, R.
TITLE	Direct Submission

JOURNAL

REMARK  
COMMENT

Submitted (16-NOV-2000) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
On Feb 14, 2001 this sequence version replaced gi:12654302.  
Contact: MGC help desk  
Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
Tissue Procurement: ATCC  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Baylor College of Medicine Human Genome  
Sequencing Center  
Center code: BCM-HGSC  
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>  
Contact: [villalon@bcm.tmc.edu](mailto:villalon@bcm.tmc.edu)  
Villalon, D.K., Luna, R.A., Hale, S.M., Hulyk, S., Lu, X., Garcia,  
A.M., Holloway, M., Telford, B., Hodgson, A., Bouck, J., Yu, W.,  
Muzny, D.M., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
 Series: IRAK Plate: 3 Row: n Column: 6  
 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 3114815.

FEATURES	SOURCE
1. <i>General</i>	
2. <i>Specific</i>	
3. <i>Other</i>	
4. <i>Other</i>	
5. <i>Other</i>	
6. <i>Other</i>	
7. <i>Other</i>	
8. <i>Other</i>	
9. <i>Other</i>	
10. <i>Other</i>	
11. <i>Other</i>	
12. <i>Other</i>	
13. <i>Other</i>	
14. <i>Other</i>	
15. <i>Other</i>	
16. <i>Other</i>	
17. <i>Other</i>	
18. <i>Other</i>	
19. <i>Other</i>	
20. <i>Other</i>	
21. <i>Other</i>	
22. <i>Other</i>	
23. <i>Other</i>	
24. <i>Other</i>	
25. <i>Other</i>	
26. <i>Other</i>	
27. <i>Other</i>	
28. <i>Other</i>	
29. <i>Other</i>	
30. <i>Other</i>	
31. <i>Other</i>	
32. <i>Other</i>	
33. <i>Other</i>	
34. <i>Other</i>	
35. <i>Other</i>	
36. <i>Other</i>	
37. <i>Other</i>	
38. <i>Other</i>	
39. <i>Other</i>	
40. <i>Other</i>	
41. <i>Other</i>	
42. <i>Other</i>	
43. <i>Other</i>	
44. <i>Other</i>	
45. <i>Other</i>	
46. <i>Other</i>	
47. <i>Other</i>	
48. <i>Other</i>	
49. <i>Other</i>	
50. <i>Other</i>	
51. <i>Other</i>	
52. <i>Other</i>	
53. <i>Other</i>	
54. <i>Other</i>	
55. <i>Other</i>	
56. <i>Other</i>	
57. <i>Other</i>	
58. <i>Other</i>	
59. <i>Other</i>	
60. <i>Other</i>	
61. <i>Other</i>	
62. <i>Other</i>	
63. <i>Other</i>	
64. <i>Other</i>	
65. <i>Other</i>	
66. <i>Other</i>	
67. <i>Other</i>	
68. <i>Other</i>	
69. <i>Other</i>	
70. <i>Other</i>	
71. <i>Other</i>	
72. <i>Other</i>	
73. <i>Other</i>	
74. <i>Other</i>	
75. <i>Other</i>	
76. <i>Other</i>	
77. <i>Other</i>	
78. <i>Other</i>	
79. <i>Other</i>	
80. <i>Other</i>	
81. <i>Other</i>	
82. <i>Other</i>	
83. <i>Other</i>	
84. <i>Other</i>	
85. <i>Other</i>	
86. <i>Other</i>	
87. <i>Other</i>	
88. <i>Other</i>	
89. <i>Other</i>	
90. <i>Other</i>	
91. <i>Other</i>	
92. <i>Other</i>	
93. <i>Other</i>	
94. <i>Other</i>	
95. <i>Other</i>	
96. <i>Other</i>	
97. <i>Other</i>	
98. <i>Other</i>	
99. <i>Other</i>	
100. <i>Other</i>	

```

i. .2280
/organism="Homo sapiens"
/db_xref="LocusID:28952"
/db_xref="taxon:9606"
/clone="MGC:5137 IMAGE:3449051"
/tissue_type="Placenta, choriocarcinoma"
/clone_lib="NIH_MGC_10"
/lab_host="DH10B"
/note="Vector: pCMV_SPORT6"
118. .2001

```

505

BASE COUNT	REA *	
450 a	712 c	699 g 409 t
ORIGIN		

Query Match 94.9%; Score 571; DB 9; Length 2280;  
Best Local Similarity 98.7%; Pred. No. 2.le-121;  
Matches 597; Conservative 0; Mismatches 5; Indels

Qy	1	tgaagaagccgcgcgttgagactgctgccgaatggagactgccaaacttgccaagctgcagc	60
tb	1271	TGAAGACCCGCGCGGTGGAGCTGCTGCCGATGGACTGCCAACCTTGCCTGAAGCTGCAGC	1330
Qy	61	tgtgtgtggaataagtgtcccaagcggatctatccacttggcgggttcagtggagagaagcacc	120
Db	1331	TTGTGGTGGAGAAATAGTGCCTCAGCGGGTCTATCCACTTTGGCGGGTCACTGGGAGAGCAAC	1390
Qy	121	agatccca--tctcgttgataccacca-ctccaaagactcagattgcagaagctgg	177
Db	1391	GGGTCCCACTCTTCGCTAGATACCGCCACCCTCCGAAAGCTCAGATTTCAGAGAGCTGG	1450
Qy	178	aattcttctcagagactgacagaataatccaaagactgcacaaagtctccggcgcgctgctg	237

```
Db 1451 AATCTTCTCGAGCGCTGCGACAGATCCAAAGAACTGCACAGAGTGTCCGGCGGCTGCTG 1510
Oy 238 aagagcccgacagaaagagaggtctataagcagctgaatcagaactgaagactctgc 297
Db 1511 AAGAGGCCCGCAGAAAGAGAGAGGTCTATAGCAGCTGATGTACAGCTGGAGACTCTGC 1570
Oy 298 ccagagatgtatcccgctgagcctacacccagcgcatccttgagatcgtggagcaatcc 357
Db 1571 CCAGAGATGTGTCCGGCTGCGCTACACCCAGCGCATCTCTGGAGATCTGTGGTAAACATCC 1630
Oy 358 ggaagcaagaagaagagatcacaagatcttgtctgatacagaagaagctcagaagaaa 417
Db 1631 GGAAGCAAGAAGAGAGATCACCAGATCTTGTCTGATACGARGAGCTTCAGAGAGAAA 1690
Oy 418 tcaactcctatctgggaagctgagcggagcgtttgcggtgactgaagcttgtattca 477
Db 1691 TCAACTCCTATCTGGGAAGCTGGACCGAGCTTTTGGGTGACTGATGAGCTTGTGTTC 1750
Oy 478 aggatgcaagaagagacatctgtctcgaagagcctataagatataagctactctgcacg 537
Db 1751 AGGATGCCAAGAAGAGCAATCTGTTCGGAAGGCCATATAAGTATCTAGCTGCTTGCACG 1810
Oy 538 agaactcgagcagcagctctccagaccatcgagacacagcagcagcagcagcagcagc 597
Db 1811 AGAAGCTGAGCAGCTATCCAGACCATCGAGGACACAGCACCATCATCGGGAGGTTTC 1870
Oy 598 gagac 602
Db 1871 GAGAC 1875

RESULT 4
AF229637 2179 bp mRNA linear ROD 22-JUN-2000
LOCUS Mus musculus bXimx40e protein (bXimx40e) mRNA, complete cds.
DEFINITION AF229637
VERSION AF229637.1 GI:7673613
KEYWORDS house mouse.
SOURCE Mus musculus
ORGANISM Euteleostomi:
Mammalia: Eutheria: Rodentia: Sciurognathi: Muridae: Murinae: Mus.
Means, G.D., Toy, D.Y., Baum, P.R. and Derry, J.M.J.
A transcript map of a 2-Mb BAC contig in the proximal portion of
the mouse X chromosome and regional mapping of the scurfy mutation
Genomics 65 (3), 213-223 (2000)
2031388
10857745
PUBMED
Derry, J.M.J.
Direct Submission
Submitted (31-JAN-2000) Molecular Biology, Immunex Corporation, 51
University St., Seattle, WA 98101, USA
FEATURES
Location/Qualifiers
1..2179
/gene="bXimx40e"
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="X"
/map="between DXMit123 and DXMit161"
1..2179
/gene="bXimx40e"
66..1952
/gene="bXimx40e"
/feature="DNA segment, Chr X, Immunex-40, expressed"
/codon_start=1
/product="bXimx40e protein"
/protein_id="AAF65951.1"
/db_xref="GI:7673614"
/translation="MEADRILIHSLFQAGTAVFVOTLPAFTLTVAVVPCLRV
INPDVGSLSHLPLPMSARFLMSLAQCHDLGYPLEGYORFLYSEPDRLPLL
FLAERLFDASEDADQADSAIFLRAIGSQIRJALPLWPPPLLRIPKVRKLOGSAL
```

BASE COUNT 452 a 612 c 629 g 446 t  
ORIGIN

Query Match 76.3% Score 459.6; DB 10; Length 2179;  
Best Local Similarity 87.3%; Pred. No. 9.7e-96;  
Matches 527; Conservative 0; Mismatches 74; Indels 3; Gaps 2;

Oy 1 tgaagagccgcgcgtgagctgctgccgatggagactgccaacacctgccaagctgcagc 60  
Db 1222 TGAAGACGACGACGCGTGAATTTGCTGCTGATGGGGCTGCCAACCTCGCCAACTGCAGC 1281  
Oy 61 ttctggtggaagatagtgccagcgggtcactccacttggcgggtcagtggaagagcacc 120  
Db 1282 TTGTGGTGGAGAGCAGCGCTCAGAGGGCTCATCTAGCAAGCCAGTGGGAGAAACACC 1341  
Oy 121 ggggtccca--tctctgtgagtlaccqcca-ctccgaaagctgcagagattgcagagagctgg 177  
Db 1342 GGGTCCCACTTCTTCTGCTGAGTACCGCCACCTCAGAGACTCCAGAGCTGTAGGAGCTGG 1401  
Oy 178 aatcttctcgagcgtgcagagagatccaaagactgcaccagagtgctcgggagctgctg 237  
Db 1402 AATGCTCTCGACGGCTGGCAGAAATCCAGGAGCTGCACACACAGTGTTCGAGCAGCTGAG 1451  
Oy 238 aagagcccgacagaaagagagagctctataagcagctgagctgcagagctgcagagctctgc 297  
Db 1452 AGGAGGCTCGCCGAGAGGAGGAGTCTATAGCAGCTGGTGTTCAGAGCTGCAGACTCTGC 1521  
Oy 298 ccagagatgtgtcccgctggcctacacccagcgcatcctcggagatcgtgggcaacatcc 357  
Db 1522 CAAAGAGATGTGTCCGGCTGCGCTATATCTCAGCGCATCTCTGGAGATTTGTGGGCAACATCC 1581  
Oy 358 gaaagcagaaagagagatcacaagatcttgtctgatacagaagagcttcagaagaaaa 417  
Db 1582 GGAAGCAGAGAGAGAGATCACTAGATCTTGTTCGGACACAAAGAACTCGAAGGAAA 1641  
Oy 418 tcaactcctatctgggaagctgagcggagcgtttgcggtgactgtagcttgattca 477  
Db 1642 TCAACTCTCTCTCTGGGAAGCTCGACCGGACATTTGCAGTCACTGATGAGCTTGTGTTC 1701  
Oy 478 aggatgcaagaagagacatcgtctcgaagagcctataagatataagctactctgcacg 537  
Db 1702 AGGATGCTAAGAGGATGATGCTGTTCGGAAGGCTTACAGTACTAGCTGCCCTGCAATG 1761  
Oy 538 aagactgcagcagctcactccagaccatcgagagacagagcaccatcagcggagaggttc 597  
Db 1762 AGAATTGCAGTCAGCTCATCCAGACCATTCAGGACACAGGCACTATCATAGGGAAGTTC 1821  
Oy 598 gaga 501  
Db 1822 GAGA 1825

RESULT 5  
BC011195 2194 bp mRNA linear ROD 30-JUL-2001  
LOCUS Mus musculus, Similar to JM1 protein, clone MGC:118697  
DEFINITION IMAGE:4188459, mRNA, complete cds.  
ACCESSION BC011195  
VERSION BC011195.1 GI:15029928  
KEYWORDS M3C.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:







```

VERSION      AF378759.1  GI:15987498
KEYWORDS
SOURCE       house mouse.
ORGANISM     Mus musculus.

REFERENCE
AUTHORS      Carlson-Walter,E.B., Watkins,D.N., Nanda,A., Vogelstein,B.,
              Kinzler,K.W. and St. Croix,B.
TITLE        Cell surface tumor endothelial markers are conserved in mice and
              humans
JOURNAL       Cancer Res. 61 (18), 6649-6655 (2001)
MEDLINE      21443268
PUBMED       11559528
REFERENCE    2 (bases 1 to 5520)
AUTHORS      Carlson-Walter,E.B., Vogelstein,B., Kinzler,K.W. and St. Croix,B.
TITLE        Direct Submission
JOURNAL       Submitted (10-MAY-2001) Oncology, Johns Hopkins University, 1650
              Orleans Street, Baltimore, MD 21231, USA
FEATURES
source       1..5520
              /organism="Mus musculus"
              /db_xref="taxon:10090"
gene         1..5520
              /gene="Tem5"
sig_peptide  127...204
              /gene="Tem5"
CDS          127..4115
              /gene="Tem5"
              /note="MTEM5: seven-pass transmembrane protein"
              /codon_start=1
              /product="tumor endothelial marker 5 precursor"
              /protein_id="AA111996.1"
              /db_xref="GI:15987499"
              /translation="MPVPPARLLLPLLLPCLLLAPGTGAPGCPVIRGCKSGERP
              KLSGGANRRVCGGGDLPDPGLPGLNGITILLNNHKTIGLRNGSFLGLSL
              LEKLDNSNITVOPCAFLGLGELKPLDGLNNRIGCLTSEIFQGLRLLRLNITSNI
              YSSLOCPVDELPAKLVDETFELTCDCLRWLLPNARNHSLQLSERTLCVPSALH
              AHALLSQELQEGALELHLYLIESLRQVYFGDRLPQCSASLYGNDIRHWHY
              NAWPESDQGLVLAENLHDCIFITSELTLSHIGWASGENECSYSTVOGNTSKKY
              EIVLETASQCPAERTNNRNGDFRNPRTLAGITAYQSCLOYPTSPVLSGGAPGTPA
              SRCDRAGRPDGYSHCLYNDITRLVYTFVLMPIANALTLAHLQRYTAEASF
              SDMMVYVQALMIQKFLGYVDQIKELVVMVMDNLVDEHLLWLAQREDKACSGI
              VGALERIGGAALSPHAOHISVNSRVALEAYLIKPHSVGLTCTAFORREVGVGAOP
              SSVGODAPVEPELADQOLPERCTIGRPNISLSFHKNSVALASIQLPSPLESLPA
              ALAPVPDCTQLQLVFNRLFRSHGNTSRGAAGPGRFGVATPVITAGISGGGV
              GNLTPEVAVSLRHWAEGADPNAAWQDGGWSSEGCRLRYSQPNVSSLYCOHLGW
              AVLMLNAFPREAGSGAGLHPVYVYPTALLLGLFTIITILNHSIHVSRGWHM
              LLNLCPHMAITSVFGVGVTLTYNQVQAVGITLHYSSLSLLMWKARVYHKELS
              WPAPTEEGEAAPGPRPMLRFYLIAGGIPLIICGITAAVNIHNRHSPVCMVWPP
              SLGAFYIPVALILPITWYELCAGLHLRSHVAONPKOGRNISLEPGEELRGSTRLSR
              GYLLNDGSLLATYSAGVTPAPPEDGDGYSPGVOLGALMIHFLYLANWACALAV
              SORWLPVRCSLYGAASALGLFVTHHCARRDRVRASNRACCPASPASASHVPARA
              LPTATEDGSPVLGGPASPASLSSGSGSGRAPPPCKLTINLOVAQSVCEASVAARGD
              EPEPTGSGSLAPRHHNLHGRVHRKRAKHRAGETGKSRKLRALRAGTSGAPFL
              LSSSEGLHNSPDSYPCSSRNSPDGLPLEGEPMLTPSEGSTSAAPIAETIRPGOR
              PSASRDLKSGSALERESKRSPPLNTISLNGAPKGGKYELASVTGAELAGSGMKT
              GLKKSQITV"
              205..4113
              /gene="Tem5"
mat_peptide  /product="tumor endothelial marker 5"
              2413..2478
              /gene="Tem5"
              /note="Region: predicted transmembrane domain"
              2518..2583
              /gene="Tem5"
              /note="Region: predicted transmembrane domain"
              2599..2664
              /gene="Tem5"
              /note="Region: predicted transmembrane domain"
              2749..2814
              /gene="Tem5"
              /note="Region: predicted transmembrane domain"

misc_feature 2853..2931
              /gene="Tem5"
              /note="Region: predicted transmembrane domain"
              3151..3207
              /gene="Tem5"
              /note="Region: predicted transmembrane domain"
              3244..3312
              /gene="Tem5"
              /note="Region: predicted transmembrane domain"
BASE COUNT  1125 a 1631 c 1600 g 1164 t
ORIGIN
Query Match 7.38; Score 42; DB 10; Length 5520;
Best Local Similarity 56.5%; Pred. No. 2;
Matches 78; Conservative 0; Mismatches 60; Indels 0; Gaps 0;
Q: 431 qgcaagattcttgagatcagccattggtacaaaaagatttttaagcttttatgtt 490
  || || || || || || || || || || || || || || || || || || || ||
Db 5383 GGGAAAGGGGAGGAGGGGATGAAGTCCAAAGAAACACATTTTAAAGCTCTGTTT 5442
  || || || || || || || || || || || || || || || || || || || ||
Q: 491 ataccatggagccatagaaagctatggaattgttttaagaactatttttaagtgtccaga 550
  || || || || || || || || || || || || || || || || || || || ||
Db 5443 ATACAATAGAAATGTTTCCAGCAGATGCTCTCTTTTAAATATATATAATCTTGCAA 5502
  || || || || || || || || || || || || || || || || || || || ||
Q: 551 cccaaaaaagaaaaaaa 568
  || || || || || || || || || || || || || || || || || || || ||
Db 5503 AAAAAAAGAAAAA 5520

RESULT 14
AC093307/c
LOCUS       AC093307          153000 bp      DNA      linear      HTG 16-AUG-2001
DEFINITION Homo sapiens chromosome 5 clone RP11-810C16, WORKING DRAFT
SEQUENCE, 13 unordered pieces.
ACCESSION   AC093307
VERSION     AC093307.1  GI:15193441
KEYWORDS    HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE      human.
ORGANISM    Homo sapiens
REFERENCE   1 (bases 1 to 153000)
AUTHORS    DOE Joint Genome Institute.
TITLE      Sequencing of Human Chromosome 5
JOURNAL    Unpublished
REFERENCE   2 (bases 1 to 153000)
AUTHORS    DOE Joint Genome Institute.
TITLE      Direct Submission
JOURNAL    Submitted (16-AUG-2001) Production Sequencing Facility, DOE Joint
              Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 1605608
Center clone name: RPC1-11_810C16
-----
Summary Statistics
Consensus quality: 142440 bases at least Q40
Consensus quality: 148623 bases at least Q30
Consensus quality: 149732 bases at least Q20
Estimated insert size: 181920; agarose-fp estimation
Estimated insert size: 151800; sum-of-contigs estimation
Quality coverage: 12.65 in Q20 bases; agarose-fp estimation
Quality coverage: 15.16 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
  * consists of 13 contigs. The true order of the pieces
  * is not known and their order in this sequence record is
  * arbitrary. Gaps between the contigs are represented as
  * runs of N, but the exact sizes of the gaps are unknown.

```

```

/note="assembly_fragment"
misc_feature 89571..101200
/note="assembly_fragment"
misc_feature 101301..117957
/note="assembly_fragment"
misc_feature 118058..133402
/note="assembly_fragment"
misc_feature 133503..150691
/note="assembly_fragment"
misc_feature 150792..151186
/note="assembly_fragment
clone_end:17
vector_side:right"
BASE COUNT 18550 a 27475 c 25972 g 47578 t 15000 others
ORIGIN
Query Match 7.4%; Score 42.4; DB 2: Length 151184;
Best Local Similarity 49.1%; Pred. No. 1.4;
Matches 112; Conservative 0; Mismatches 116; Indels 0; Gaps 0;
Oy 347 aaatagccagctaatagccagaaatcatgaccctgaagagagatgaaatttcaagtg 405
Db 15206 AATTAAGCAGCTAGGTTTCTATATCAATGGAATAGATTAATCTAAATGTCAACGAG 15265
Oy 407 tgatccaggcagcagctccagtgatggcaaggtctctgagaatcagcatttgatgaaa 466
Db 15266 ACATACATGGAATGATGATTTTAAACAAGGTGTCATCAGAGTAACCTCAATGGGAGAACGA 15325
Oy 467 aaatattttaagcttttattgtataccatggagccatgagaagcctatgatttta 526
Db 15326 AAAATTITCCACAAATGTGCCAGAACAACTGGATATCATCTTTGGGAGGAACAATGGA 15385
Oy 527 agaatatttaagctttccagcccaaggaaggaaggaaggaaggaaggaaggaaggaagga 574
Db 15386 AGTACTATTTTTCCTCAGCTGATACCAAAAAAATAAAAAAATAAAAAAATAAAAAA 15433
RESULT 12
AC022421/c
LOCUS Homo sapiens chromosome 5 clone CTD-2079J19, WORKING DRAFT
DEFINITION AC022421
ACCESSION AC022421
VERSION AC022421.6 GI:13699635
KEYWORDS HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Eumetazoa;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 110146)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 5
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 110146)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (03-FEB-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Michell Drive, Walnut Creek, CA 94598, USA
COMMENT On Apr 20, 2001 this sequence version replaced gi:1030553.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 658120
Center clone name: CITB-H1_2079J19
-----
Summary Statistics
Consensus quality: 92776 bases at least Q40
Consensus quality: 98775 bases at least Q30
Consensus quality: 101176 bases at least Q20

```

Qy 527 agaactattttaaagtgttcagaccacaaagagagaaaaa 574  
 || ||||| | | | | | | | | | | | | | | | | | |  
 Db 15386 AGTACTATTTTTCCTCAGTCGTACCAAAAAAATAA 15433

## RESULT 11

```

AC068750      151186 bp   DNA      linear      HTG 05-SEP-2000
LOCUS      Homo sapiens chromosome 11 clone RP11-63B15 map 11, WORKING DRAFT
DEFINITION      SEQUENCE, 17 unordered pieces.
AC068750
ACCESSION      AC068750
VERSION        3
KEYWORDS      HTG: HIGS-PHASE1; HIGS-DRAFT.
SOURCE        human.
ORGANISM      Homo sapiens
               Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
               1 (bases 1 to 151186)
               Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
               Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
               Boguslavskiy,L., Boukhaltier,B., Brown,A., Burkett,G.,
               Campopiano,A., Castelle,A., Choepel,Y., Colangelo,W., Collins,S.,
               Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
               Galagan,J., Gardyna,S., Glinde,S., Goyette,P., FitzHugh,W., Gage,D.,
               Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
               Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
               Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczyk,J.,
               Levine,R., Liu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
               McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
               Meldrum,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morris,J.,
               Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
               O'Neill,D., Ollivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
               Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
               Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
               Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
               Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigglio,J.,
               Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
               Young,G., Zainoun,J., Zimmer,A. and Zody,M.
               Direct Submission
               Submitted (08-MAY-2000) Whitehead Institute/MIT Center for Genome
               Research, 320 Charles Street, Cambridge, MA 02141, USA
               On Sep 5, 2000 this sequence version replaced gi:8247941.
               All repeats were identified using RepeatMasker:
               Smit, A.F.A. & Green, P. (1996-1997)
               http://ftp.genome.washington.edu/RM/RepeatMasker.html
               ----- Genome Center
               Center: Whitehead Institute/ MIT Center for Genome Research
               Center code: WIBR
               Web site: http://www.seq.wi.mit.edu
               Contact: sequence_submissions@genome.wi.mit.edu
               ----- Project Information
               Center project name: L9968
               Center clone name: 63.B.15
               ----- Summary Statistics
               Sequencing vector: M13; M77815; 95% of reads
               Sequencing vector: Plasmid; n/a; 0.0% of reads
               4.54873645209386Chemistry: Dye-terminator Big Dye; 100% of
               reads
               Assembly program: Phrap; version 0.960731
               Consensus quality: 143431 bases at least Q40
               Consensus quality: 146986 bases at least Q30
               Consensus quality: 148556 bases at least Q20
               Insert size: 151000; agarose-fp
               Insert size: 149586; sum-of-contigs
               Quality coverage: 4.8 in Q20 bases; agarose-fp
               Quality cov.
               * NOTE: This is a 'working draft' sequence. It currently

```

```

* consists of 17 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 16641: contig of 16641 bp in length
* 16642 16741: gap of 100 bp
* 15742 36912: contig of 20171 bp in length
* 36913 37012: gap of 100 bp
* 37013 38128: contig of 1116 bp in length
* 38129 38228: gap of 100 bp
* 38229 39683: contig of 1455 bp in length
* 39684 39783: gap of 100 bp
* 39784 42531: contig of 2748 bp in length
* 42532 42631: gap of 100 bp
* 42632 45365: contig of 2734 bp in length
* 45366 45465: gap of 100 bp
* 45466 48220: contig of 2755 bp in length
* 48221 48320: gap of 100 bp
* 48321 53484: contig of 5164 bp in length
* 53485 53584: gap of 100 bp
* 53585 62205: contig of 8621 bp in length
* 62206 62305: gap of 100 bp
* 62306 68614: contig of 6309 bp in length
* 68615 68714: gap of 100 bp
* 68715 79766: contig of 11052 bp in length
* 79767 79866: gap of 100 bp
* 79867 89570: contig of 9704 bp in length
* 89571 89670: gap of 100 bp
* 89671 101200: contig of 11530 bp in length
* 101201 101300: gap of 100 bp
* 101301 117957: contig of 16657 bp in length
* 117958 118057: gap of 100 bp
* 118058 133402: contig of 15345 bp in length
* 133403 133502: gap of 100 bp
* 133503 150691: contig of 17189 bp in length
* 150692 150791: gap of 100 bp
* 150792 151186: contig of 395 bp in length.
FEATURES
    source          1..151186
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /chromosome="11"
                    /map="11"
                    /clone="RP11-63B15"
                    /clone_lib="RPC1-11 Human Male BAC"
    misc_feature    1..16641
                    /note="assembly_fragment"
                    clone_end:SP6
                    vector_side:left
                    16742..36912
                    /note="assembly_fragment"
                    37013..38128
                    /note="assembly_fragment"
                    38229..39683
                    /note="assembly_fragment"
                    39784..42531
                    /note="assembly_fragment"
                    42532..45365
                    /note="assembly_fragment"
                    45466..48220
                    /note="assembly_fragment"
                    48321..53484
                    /note="assembly_fragment"
                    53585..62205
                    /note="assembly_fragment"
                    62306..68614
                    /note="assembly_fragment"
                    68715..79766
                    /note="assembly_fragment"
                    79867..89570
                    /note="assembly_fragment"

```





```

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Lodes,M.J., Wang,T., Mohamath,R. and Indirias,C.Y.
TITLE Compositions and methods for the therapy and diagnosis of lung
cancer
JOURNAL Patent: WO 0177158-A 245 18-OCT-2001;
CORIXA CORPORATION (US)
FEATURES
source 1..473
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 144 a 105 c 84 g 140 t
ORIGIN
Query Match 54.3%; Score 370; DB 6; Length 473;
Best Local Similarity 99.5%; Pred. No. 1.8e-84;
Matches 392; Conservative 0; Mismatches 0; Indels 2; Gaps 2;

Oy 1 acttgataaaagagatccatgaatgaattttatactgcattcctttacattagcca 60
|||||
Db 82 ACTTGATATAAAAGGATATCCATAAAGAAATATTTATATCTGCAATCCCTTACATTAGCCA 141
|||||

Oy 61 ctaaatcagttattgcttgatgaagaccctttcacagaatcctatgattgagccatttca 120
|||||
Db 142 CTAATACGTTATGCTTGATGAAGACCTTTTCAAGAACTTATGATTCAGCATTTCA 201
|||||

Oy 121 cttagctacttcataccatgccttaaaagaggcaggtttctcaaaagcagaaacatgcc 180
|||||
Db 202 CTGGCTACTTCATACCCATGCCCTTAAAGAGGGCGAGTTTCTCAAAAGCAGAAACATGCC 261
|||||

Oy 181 gccagttctcaagttttctccttaactccatttgatgtgaaggcagctgcccccaatg 240
|||||
Db 262 GCCAGTTCTCAAGTTTTCCTCCTTAACCTCCATTTGAATGTAAGGCGAGCTGGCCCCCAATG 321
|||||

Oy 241 tggaggagtcggaacattttcctgaattccattttctgttcggtcgttaaatgacagttt 300
|||||
Db 322 TGGGAGGTCGGAACATTTTCTGAATTCCTCAATTTCTTGTTTCGGCGCTAANTGACAGTTT 381
|||||

Oy 301 ctgtcattacttagattcccgatctttcccaaaagtgattgatttacaagagccagcta 360
|||||
Db 382 CTGTCAITACTTACATT-CCGATCTTCCCAAGGCTGTGATTTTACAAGAGGCCAGCTA 440
|||||

Oy 361 atagccagaatcatgacctgaaagagagatga 394
|||||
Db 441 ATAG-CAGAAATCATGACCCCTGAAGAGAGATGA 473
|||||

RESULT 7
AX321385 LOCUS AX321385 473 bp DNA linear PAT 15-DEC-2001
DEFINITION Sequence 402 from Patent WO0177158.
ACCESSION AX321385
VERSION AX321385.1 GI:17905282
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Lodes,M.J., Wang,T., Mohamath,R. and Indirias,C.Y.
TITLE Compositions and methods for the therapy and diagnosis of lung
cancer
JOURNAL Patent: WO 0177158-A 402 18-OCT-2001;
CORIXA CORPORATION (US)
FEATURES
source 1..473
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 144 a 105 c 84 g 140 t
ORIGIN

```

```

Query Match 64.3%; Score 370; DB 6; Length 473;
Best Local Similarity 99.5%; Pred. No. 1.8e-84;
Matches 392; Conservative 0; Mismatches 0; Indels 2; Gaps 2;

Oy 1 acttgataaaagagatccatgaatgaattttatactgcattcctttacattagcca 60
|||||
Db 82 ACTTGATATAAAAGGATATCCATAAAGAAATATTTATATCTGCAATCCCTTACATTAGCCA 141
|||||

Oy 61 ctaaatcagttattgcttgatgaagaccctttcacagaatcctatgattgagccatttca 120
|||||
Db 142 CTAATACGTTATGCTTGATGAAGACCTTTTCAAGAACTTATGATTCAGCATTTCA 201
|||||

Oy 121 cttagctacttcataccatgccttaaaagaggcaggtttctcaaaagcagaaacatgcc 180
|||||
Db 202 CTGGCTACTTCATACCCATGCCCTTAAAGAGGGCGAGTTTCTCAAAAGCAGAAACATGCC 261
|||||

Oy 181 gccagttctcaagttttctccttaactccatttgatgtgaaggcagctgcccccaatg 240
|||||
Db 262 GCCAGTTCTCAAGTTTTCCTCCTTAACCTCCATTTGAATGTAAGGCGAGCTGGCCCCCAATG 321
|||||

Oy 241 tggaggagtcggaacattttcctgaattccattttctgttcggtcgttaaatgacagttt 300
|||||
Db 322 TGGGAGGTCGGAACATTTTCTGAATTCCTCAATTTCTTGTTTCGGCGCTAANTGACAGTTT 381
|||||

Oy 301 ctgtcattacttagattcccgatctttcccaaaagtgattgatttacaagagccagcta 360
|||||
Db 382 CTGTCAITACTTACATT-CCGATCTTCCCAAGGCTGTGATTTTACAAGAGGCCAGCTA 440
|||||

Oy 361 atagccagaatcatgacctgaaagagagatga 394
|||||
Db 441 ATAG-CAGAAATCATGACCCCTGAAGAGAGATGA 473
|||||

RESULT 8
AX335247/c LOCUS AX335247 294 bp DNA linear PAT 09-JAN-2002
DEFINITION Sequence 5756 from Patent WO0194629.
ACCESSION AX335247
VERSION AX335247.1 GI:18125966
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Young,P.E., Augustus,M., Carter,K.C., Ebner,R., Endress,G.,
Harrigan,S., Soppet,D.R. and Weaver,Z.
TITLE Cancer gene determination and therapeutic screening using signature
gene sets.
JOURNAL Patent: WO 0194629-A 5756 13-DEC-2001;
Avalon Pharmaceuticals (US)
FEATURES
source 1..294
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 85 a 62 c 51 g 96 t
ORIGIN

```

```

Query Match 43.5%; Score 250; DB 6; Length 294;
Best Local Similarity 98.7%; Pred. No. 1e-53;
Matches 294; Conservative 0; Mismatches 0; Indels 4; Gaps 4;

Oy 264 aattcccatcttctgttcgcgcgctaaagacagtttctgtcattacttagattcccgat 323
|||||
Db 294 AATTCCCATTTTCTGTTCTCGCGCTAATGACAGTTTCTGTCTACTTAGATT-CCGAT 236
|||||

Oy 324 ctctcccaaaaggtgttgatttacaagagccagctgaatagccagaataatcatgacctga 383
|||||
Db 235 CTCTCCCAAAAGGTGTGATTTACAAGAGGCCAGCTAATAG-CAGAAATCATGACCCCTGA 177
|||||

```



Oy 541 gtgttcagaccacaaagagaaa 563  
|||||  
Db 768 GTGTTCAGACCCAAAAA 790

## RESULT 4

LOCUS ARI44162 654 bp DNA linear PAT 08-AUG-2001  
DEFINITION Sequence 27 from patent US 6210883.  
ACCESSION ARI44162  
VERSION ARI44162.1 GI:15105029  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 554)  
AUTHORS Reed, S.G. and Wang, T. Tong.  
TITLE Compounds and methods for diagnosis of lung cancer  
JOURNAL Patent: US 6210883-A 27 03-APR-2001;  
FEATURES Location/Qualifiers  
source  
1. .654  
/organism="unknown"  
BASE COUNT 218 a 130 c 119 g 180 t 7 others  
ORIGIN

Query Match 79.3%; Score 455.8; DB 6; Length 654;  
Best Local Similarity 94.6%; Pred. No. 1.8e-106;  
Matches 543; Conservative 0; Mismatches 24; Indels 7; Gaps 7;

Oy 1 acttgatataaaagagatccataatgaatattttatctactgcattcttacattagcca 60  
|||||  
Db 88 ACTTGATATAAAAGGATAICCATATGAATATTTTATCTGCATCTTTACATTAGCCA 147  
Oy 61 ctaataacgttattgcttgatgaagaccctttcacagaatcctatgattgcagcatttca 120  
|||||  
Db 148 CTAATAACGTTATTGCTTGATGAAGACCTTTCACAGAACTCTATGTGATTGCAGCATTTCA 207  
Oy 121 ctggctacttcataccatgctttaaagggcgagtttctcaaaagcagaacaatgcc 180  
|||||  
Db 208 CTGGCTACTTCATACCCATGCTTAAAGAGGGCGAGTTCTCAAAAGCAGAAACATGCC 257  
Oy 181 gccagttctcaagtttctcctcaactccatttgaatgtaaggcgagctggccccaatg 240  
|||||  
Db 268 GCCAGTTCTCAAGTTTTCCTCTCACTCACTCCATTGTAATGTAAGGCGCAGCTGGCCCCCAATG 327  
Oy 241 tggggaggctccgaacatttctgaaattcccatlcttctgctcgcggcctaataagcagttt 300  
|||||  
Db 328 TGGGGAGGTCGGAACATTTCTGAATTCCTCATTTCTGTTCGCGGCTAAATGACAGTTT 387  
Oy 301 ctgtcattacttagatcccgatcttcccaagggtgttgatttacaagagggccagcta 360  
|||||  
Db 388 CTGTCTATTACTTAGATT-CCGATCTTTCCCAAAGGTGTGATTTCACAAAGAGGCCAGCTA 446  
Oy 361 atagccagaatacatgacctgaagagagatgaatttcaagctgtgagccagcagga 420  
|||||  
Db 447 ATAG-CAGAAATCATGACCTTGAAGAGAGATGAAA-TTCAAGCTGTGAGCCAGGCAGGA 504  
Oy 421 gctccagatggcaagggttcttgaatcagccatttggatcacaagaaagatttttaag 480  
|||||  
Db 505 NCT-CAGTATGGCAAGG-TCTTGAGAAATCGCCATTTGGTACAAAAA-AATTTTAAAG 561  
Oy 481 cttttatgttataccatgagccatgagaaggctatgattgtttaagaactattttaa 540  
|||||  
Db 562 CNITATGTTTATACCATGGAACCATAGAAANGCAAGGGAATTTGTTAAGANAATTTTAA 521  
Oy 541 gtgttcagaccacaaagagaaaaa 574  
|||||  
Db 622 GTG-TCCAGACCCCAANGAANGAAAAA 554

## RESULT 5

ARI176349  
LOCUS ARI176349 654 bp DNA linear PAT 17-DEC-2001  
DEFINITION Sequence 27 from patent US 6312695.  
ACCESSION ARI176349  
VERSION ARI176349.1 GI:17918704  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 654)  
AUTHORS Reed, S.G. and Wang, T. Tong.  
TITLE Compounds and methods for therapy of lung cancer  
JOURNAL Patent: US 6312695-A 27 06-NOV-2001;  
FEATURES Location/Qualifiers  
source  
1. .654  
/organism="unknown"  
BASE COUNT 218 a 130 c 119 g 180 t 7 others  
ORIGIN

Query Match 79.3%; Score 455.8; DB 6; Length 654;  
Best Local Similarity 94.6%; Pred. No. 1.8e-106;  
Matches 543; Conservative 0; Mismatches 24; Indels 7; Gaps 7;

Oy 1 acttgatataaaagagatccataatgaatattttatctactgcattcttacattagcca 60  
|||||  
Db 88 ACTTGATATAAAAGGATAICCATATGAATATTTTATCTGCATCTTTACATTAGCCA 147  
Oy 61 ctaataacgttattgcttgatgaagaccctttcacagaatcctatgattgcagcatttca 120  
|||||  
Db 148 CTAATAACGTTATTGCTTGATGAAGACCTTTCACAGAACTCTATGTGATTGCAGCATTTCA 207  
Oy 121 ctggctacttcataccatgctttaaagggcgagtttctcaaaagcagaacaatgcc 180  
|||||  
Db 208 CTGGCTACTTCATACCCATGCTTAAAGAGGGCGAGTTCTCAAAAGCAGAAACATGCC 267  
Oy 181 gccagttctcaagtttctcctcaactccatttgaatgtaaggcgagctggccccaatg 240  
|||||  
Db 268 GCCAGTTCTCAAGTTTTCCTCTCACTCACTCCATTGTAATGTAAGGCGCAGCTGGCCCCCAATG 327  
Oy 241 tggggaggctccgaacatttctgaaattcccatlcttctgctcgcggcctaataagcagttt 300  
|||||  
Db 328 TGGGGAGGTCGGAACATTTCTGAATTCCTCATTTCTGTTCGCGGCTAAATGACAGTTT 387  
Oy 301 ctgtcattacttagatcccgatcttcccaagggtgttgatttacaagagggccagcta 360  
|||||  
Db 388 CTGTCTATTACTTAGATT-CCGATCTTTCCCAAAGGTGTGATTTCACAAAGAGGCCAGCTA 446  
Oy 361 atagccagaatacatgacctgaagagagatgaatttcaagctgtgagccagcagga 420  
|||||  
Db 447 ATAG-CAGAAATCATGACCTTGAAGAGAGATGAAA-TTCAAGCTGTGAGCCAGGCAGGA 504  
Oy 421 gctccagatggcaagggttcttgaatcagccatttggatcacaagaaagatttttaag 480  
|||||  
Db 505 NCT-CAGTATGGCAAGG-TCTTGAGAAATCGCCATTTGGTACAAAAA-AATTTTAAAG 561  
Oy 481 cttttatgttataccatgagccatgagaaggctatgattgtttaagaactattttaa 540  
|||||  
Db 562 CNITATGTTTATACCATGGAACCATAGAAANGCAAGGGAATTTGTTAAGANAATTTTAA 621  
Oy 541 gtgttcagaccacaaagagaaaaa 574  
|||||  
Db 622 GTG-TCCAGACCCCAANGAANGAAAAA 654

## RESULT 6

AX321228  
LOCUS AX321228 473 bp DNA linear PAT 15-DEC-2001  
DEFINITION Sequence 245 from Patent WO0177168.  
ACCESSION AX321228  
VERSION AX321228.1 GI:17904842  
KEYWORDS human.

BASE COUNT 51905 a 36985 c 38442 g 56185 t 100 others  
ORIGIN

Query Match 91.4%; Score 525.4; DB 2: Length 183617;  
Best Local Similarity 99.1%; Pred. No. 2.2e-124;  
Matches 570; Conservative 0; Mismatches 1; Indels 4; Gaps 4;  
  
Oy 1 acttgatataaaagagatccataatgaattttatctgcatctttacattagcca 60  
|||||  
Db 32567 ACTTGATATAAAAGGATATCCATAATGAATATTTATACGATCTTTACATAGCCA 32624  
Oy 61 ctataatcgctatgttgatgaagaccctttccaaatctctatgcatgacattca 120  
|||||  
Db 32627 CTAATAACGTTATGCTTGAAGAACCTTTCACAGAACTCTATGTAATGACATTTCA 32686  
Oy 121 ctggctactctataccatgccttaaaagggcgagtttctcaaaagcagaaacatgcc 180  
|||||  
Db 32687 CTGGCTACTTCATACCATGCGCTTAAGAGGGGCGAGTTTTCACAAAGCAGAAACATGCC 32745  
Oy 181 gccagttctcaagtttctctcttaactccattggaatgaagggcgagctggcccccaatg 240  
|||||  
Db 32747 GCCAGTTCTCAAGTTTCTCTCTTAACCTCCATTTGAATGTAAGGCGAGCTGGCCCCCAATG 32806  
Oy 241 tggggaggtccgaacatttctgaattccatttctgttcgagctaaatgacagttt 300  
|||||  
Db 32807 TGGGAGGTCGGAACATTTTCTGAATTCCTATTTCTTGTCGCGCTTAATGACAGTTT 32866  
Oy 301 ctgtcattacttagattcccgatctttcccaaaaggttgatttacaagagggccagcta 360  
|||||  
Db 32867 CTGTCTATTACTTAGATT-CCGATCTTTCCTCAAGGTTGTTGATTACAAAGAGGCCAGCTA 32925  
Oy 361 atagccagaatcatgaacctgaagagagataaatttcaagctatgaagcagcagga 420  
|||||  
Db 32926 ATAG-CAGAAATCATGACCTGGAAGAGAGATGAAA-TTCAAGCTGTGAGCCAGGACGGA 32983  
Oy 421 gctccagtggaagaggttcttgagaatcagccatttggtacaaaagagatttttaag 480  
|||||  
Db 32984 GCT-CAGTATGGCAAGGTTCTTGAGAAATCAGCCATTGTTGTACAAAAGATTTTAAAG 33042  
Oy 481 cttttatgttatccatgagacatagaagagctatgagttgtttaagaactttttaa 540  
|||||  
Db 33043 CTTTATGTTATACCATGGAGCCATAGAAGGCTATGGATGTTTAAAGAACTATTTTAA 33102  
Oy 541 gtgtccagagcccaaaagaaaaaataaaaaa 575  
|||||  
Db 33103 GTGTTCAGACCCCAAAAGCAAAATATAAAAAA 33137

RESULT 3  
HUMY291G09  
LOCUS HUMY291G09 790 bp mRNA linear PRI 29-AUG-1998  
DEFINITION Homo sapiens full length insert cDNA clone y291G09.  
ACCESSION AF086101  
VERSION AF086101.1 GI:3483446  
KEYWORDS FLI-CDNA.  
SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 790)  
REFERENCE Woessner,J., Tan,F., Marra,M., Kucaba,T., Yandell,M., Martin,J.,  
Marth,G., Bowles,L., Wylie,T., Boxers,Y., Sleptoe,M., Theising,R.,  
Geisel,S., Allen,M., Underwood,K., Chappell,J., Person,B.,  
Gibbons,K., Harvey,N., Pape,D., Chamberlain,A., Morales,R.,  
Schurk,R., Pitter,E., Kohn,S., Swaller,I., Behymer,K., Hillier,L.,  
Wilson,R. and Waterston,R.  
Full Clone Sequencing of the Longest Available Member from Each  
Unigene Cluster

JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 790)  
AUTHORS Waterston,R.

TITLE  
JOURNAL  
COMMENT

Direct Submission  
Submitted (24-AUG-1998) Department of Genetics, Washington  
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
SUBMITTED BY:  
Genome Sequencing Center  
Department of Genetics  
Washington University  
St. Louis MO 63108, USA  
http://genome.wustl.edu/gsc  
mailto:est@watson.wustl.edu

NOTICE: This sequence represents the full insert of this cDNA. No  
attempt has been made to verify whether this corresponds to the  
full-length of the original mRNA from which it was derived. We  
have tried to obtain double-stranded, or double chemistry sequence  
across the entire clone, but potentially, there are areas in the  
sequence where this level of coverage was not achieved.  
Nevertheless, we are confident of the accuracy of this sequence as  
all regions of low quality, as defined by PHRAP (P. Green, in  
preparation), were visually inspected and edited accordingly. The  
consensus quality values for this sequence have been submitted  
separately.

The location of this clone is unknown.

FEATURES  
Source

Location/Qualifiers  
1..790  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:290464"  
/clone\_lib="Soares\_multiple\_sclerosis\_2NDHMS"

BASE COUNT 255 a 152 c 145 g 238 t  
ORIGIN

Query Match 89.0%; Score 511.8; DB 9: Length 790;  
Best Local Similarity 98.9%; Pred. No. 8e-121;  
Matches 557; Conservative 0; Mismatches 2; Indels 4; Gaps 4;  
  
Oy 1 acttgatataaaagagatccataatgaattttatctgcatctttacattagcca 60  
|||||  
Db 232 ACTTGATATAAAAGGATATCCATAATGAATATTTATACGATCTTTACATAGCCA 291  
Oy 61 ctataatcgctatgttgatgaagaccctttcacagaatccctatgagcagattcca 120  
|||||  
Db 292 CTAATAACGTTATGCTTGAAGAACCTTTCACAGAACTCTATGATGTCAGCATTTCA 351  
Oy 121 ctggctactctataccatgccttaaaagggcgagtttctcaaaagcagaaacatgcc 180  
|||||  
Db 352 CTGGCTACTTCATACCCATGCCCTTAAGAGGGCGAGTTTCTCAAAAGCAGAAACATGCC 411  
Oy 181 gccagttctcaagtttctctcttaactccattggaatgaagggcgagctggcccccaatg 240  
|||||  
Db 412 GCCAGTTCTCAAGTTTCTCTCTTAACCTCCATTTGAATGTAAGGCGAGCTGGCCCCCAATG 471  
Oy 241 tggggaggtccgaacatttctgaaattcccatllcttcttcgcgagctaaatgacagttt 300  
|||||  
Db 472 TGGGAGGTCGGAACATTTTCTGAATTCCTATTTCTTGTCGCGCTAAATGACAGATT 531  
Oy 301 ctgtcattacttagattcccgatctttcccaaaaggttgatttacaagagggccagcta 360  
|||||  
Db 532 CTGTCAITACTTAGATT-CCGATCTTTCCTCAAGGTTGTTGATTTCACAAAGAGGCCAGCTA 590  
Oy 361 atagccagaatcatgaacctgaagagagataaatttcaagctatgaagcagcagga 420  
|||||  
Db 591 ATAG-CAGAAATCATGACCTGGAAGAGAGATGAAA-TTCAAGCTGTGAGCCAGGACGGA 648  
Oy 421 gctccagtggaagaggttcttgagaatcagccatttggtacaaaagagatttttaag 480  
|||||  
Db 649 GCT-CAGTATGGCAAGGTTCTTGAGAAATCAGCCATTGTTGTACAAAAGATTTTAAAG 707  
Oy 481 cttttatgttatccatgagacatagaagagctatgagttgtttaagaactttttaa 540  
|||||  
Db 708 CTTTATGTTATACCATGGAGCCCTAAGAAAGCTATGGATGTTTAAAGAACTATTTTAA 767

```

/note="L2 repeat: matches 2580. 2248 of consensus"
42426..42626
/note="L2 repeat: matches 2023. 2228 of consensus"
43765..43899
/note="MULTI1 repeat: matches 59. 205 of consensus"
43997..45345
/note="LIPAB repeat: matches 4813. 6161 of consensus"
45917..46305
/note="match: GSS: Em:A0075236"
46515..46785
/note="L2 repeat: matches 2408. 2468 of consensus"
47901..48024
/note="AluJb repeat: matches 3. 125 of consensus"
49747..49910
/note="MIR repeat: matches 56. 244 of consensus"
50423..50675
/note="LIMB3 repeat: matches 5926. 6176 of consensus"
51366..51550
/note="AluJ repeat: matches 127. 131 of consensus"
53273..53576
/note="AluSx repeat: matches 3. 110 of consensus"
54377..54432
/note="MIR repeat: matches 53. 111 of consensus"
54496..54612
/note="MIR repeat: matches 61. 184 of consensus"
54927..55102
/note="MIR repeat: matches 3. 183 of consensus"
55627..55690
/note="2 copies 32 mer 92% conserved"
complement(55692..55879)
/note="match: STS: Em:L30443"
complement(55692..55873)
/note="match: STS: Em:L18003"

Query Match 91.4%; Score 525.4; DB 9; Length 182914;
Best Local Similarity 95.18; Pred. No. 2.2e-124;
Matches 570; Conservative 0; Mismatches 1; Indels 4; Gaps 4;

Oy 1 actgatataaaagatatacataatgaatatttttatactgcatctttacattagccca 50
Db 173021 ACTGATATAAAAGGATATCATATGAATATTTTATCTGCAATCTTTACATTAGCCA 173080

Oy 61 ctaataacatttacttgatgaagaacaccttcacagaactctatgatttcagcatttca 120
Db 173081 CTAATAACGATTAATGCTTGATGAAGACCTTTTCAGAAATCTTATG:ATTCAGGCAATTCA 173140

Oy 121 ctggctacttcataccatgccttaagaggggcagtttctcacaagcagaacaatgcc 180
Db 173141 CTGGCTACTTCATACCAATGCTTAAGAGGGGCGAGTTTCTCAAAAGCAGAAACATGCC 173200

Oy 181 gccagttctcaagtttctctactcactccatttgatgaatgaaggcagctgaccccaatg 240
Db 173201 GCCAGTCTCAAGTTTCTCTACTCTCATTTGAATGAAGGGGAGCTGTGGCCGCCCAATG 173260

Oy 241 tggggagggtccgaacatttcttgaatttcccatttcttcttgcgcctlaaatgacagttt 300
Db 173261 TGGGGAGGTCCGNACATTTTCGTAATCCCATTTCTTGTCGGCTTAATGACAGTTT 173320

Oy 301 ctctactacttagattcccgatcttcccaagtgattgatttcacaagagcgcaacta 360
Db 173321 CTCTCAATTACTTAGATT-CCGATCTTCTCCAAAGGTGTTGATTTCACAAAGAGGCGCAGCTA 173379

Oy 361 atagccagaataatcatgacctgaagagagatgaatttcaagcctagccaggcagaa 420
Db 173380 ATAG-CAGAAATCATGACCTTGAAGAGAGATGANA-TTCAGGCTGTAGCCAGGCAAGA 173437

Oy 421 gctccagatggcaaggttcttgagaatcagccatttggttcacaaagagattttanaag 480
Db 173438 GCT-CAGATGGCAAGGTTCTTGAGAAATCAGCCATTGTTGTTACAAAGAGATTTTAAAG 173495

Oy 481 cttttatgttataccatgagccatgaaagcctatggtatgcttttgaacactatttttaa 540
Db 173497 CTTTATGTTATACCATGGAGCCATAGAAAGGCTATGGATTGTTTAAAGACTATTTTANA 173556
```

```
Oy 541 gtattccagaccacaaagagaaaaaataaaaaa 575
|||||
Db 173557 GTGTTCCAGACCCAAAAGGAAAATAAANA 173591
```

RESULT 2

AC007524  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

183617 bp DNA linear HTG 14-OCT-2001  
Homo sapiens chromosome 9 clone RP11-30A1 map 9, \*\*\* SEQUENCING IN  
PROGRESS \*\*\* 2 ordered pieces.  
AC007524.12 G1:16118190  
HTG: HTGS-PHASE2; HTGS-FULLTOP; HTGS-ACTIVEFIN.  
human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS

1 (bases 1 to 183617)  
Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
Homo sapiens chromosome 9, clone RP11-30A1  
Unpublished  
2 (bases 1 to 183617)  
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,  
Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,  
Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,  
Cooke,P., DeArelano,K., Depayre,E., Devon,K., Dekar,K.,  
Donelan,L., Doyle,M., Ferreira,P., Fitzhugh,W., Forrest,C.,  
Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,  
Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,  
Karatas,A., Lehoczy,J., Lieu,C., Locke,K., Macdonald,P.,  
Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,  
Meldrum,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,  
Naylor,J., Niloff,M., O'Connor,I., O'Donnell,P., Pavlin,B.,  
Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,  
Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,  
Testaye,S., Torruella-Miller,I., Vassiliev,H., Vo.A., Wagner,A.,  
Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.

TITLE  
JOURNAL  
COMMENT

Submitted (11-MAY-1999) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Oct 14, 2001 this sequence version replaced gi:14579797.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: http://www.seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu  
----- Project Information  
Center project name: L602  
Center clone name: 30\_A\_1  
-----

NOTE: This is a 'working draft' sequence. It currently  
consists of 2 contigs. Gaps between the contigs  
are represented as runs of N. The order of the pieces  
is believed to be correct as given, however the sizes  
of the gaps between them are based on estimates that have  
been provided by the submitter.  
This sequence will be replaced  
by the finished sequence as soon as it is available and  
the accession number will be preserved.  
1 47086: contig of 47086 bp in length  
47087 47186: gap of 100 bp  
47187 183617: contig of 136431 bp in length.  
Location/Qualifiers  
1. 183617  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="9"  
/map="9"  
/clone="RP11-30A1"

FEATURES  
Source



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM - nucleic - nucleic search, using sw model

Run on: April 29, 2002, 20:03:59 ; Search time 8020.64 Seconds  
(without alignments)  
1500.224 Million cell updates/sec

Title: US-09-248-178-65

Perfect score: 575

Sequence: 1 actgataataaaaggatat.....aaaggaataaaataaaaaa 575

Scoring table:

IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Sea-ched: \* 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Data base :

GenEmbl: \*

1: gb\_ba:\*

2: gb\_htg:\*

3: gb\_in:\*

4: gb\_om:\*

5: gb\_ov:\*

6: gb\_pat:\*

7: gb\_ph:\*

8: gb\_pl:\*

9: gb\_pr:\*

10: gb\_ro:\*

11: gb\_sts:\*

12: gb\_sy:\*

13: gb\_un:\*

14: gb\_vi:\*

15: em\_ba:\*

16: em\_fun:\*

17: em\_hum:\*

18: em\_in:\*

19: em\_mu:\*

20: em\_om:\*

21: em\_or:\*

22: em\_ov:\*

23: em\_pat:\*

24: em\_ph:\*

25: em\_pl:\*

26: em\_ro:\*

27: em\_sts:\*

28: em\_un:\*

29: em\_vi:\*

30: em\_htg\_hum:\*

31: em\_htg\_inv:\*

32: em\_htg\_other:\*

33: em\_htg\_inv:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
-----					

1	525.4	91.4	182914	9	AL390777	Human DNA
2	525.4	91.4	183617	2	AC007524	Homo sapi
3	511.8	89.0	790	9	HUMY291G09	Homo sapi
4	455.8	79.3	654	6	AR144162	Sequence
5	455.8	79.3	654	6	AR176349	Sequence
6	370	64.3	473	6	AX321228	Sequence
7	370	64.3	473	6	AX321385	Sequence
8	250	43.5	294	6	AX335247	Sequence
9	43	7.5	148824	2	AC025640	Homo sapi
10	42.4	7.4	143104	9	AL606503	Human DNA
11	42.4	7.4	151186	2	AC068750	Homo sapi
12	42.2	7.3	110146	2	AC022421	Homo sapi
13	42	7.3	5520	10	AF378759	Mus muscu
14	42	7.3	153000	2	AC093307	Homo sapi
15	42	7.3	173294	2	AC109447	Homo sapi
16	41.4	7.2	160666	2	AC105755	Homo sapi
17	41.4	7.2	168447	2	AL671985	Homo sapi
18	41.4	7.2	171962	2	AC013537	Homo sapi
19	41.4	7.2	175488	2	AC013556	Homo sapi
20	41.2	7.2	199551	2	AC006281	Plasmodiu
21	41.2	7.2	204182	2	AL671870	Mus muscu
22	41	7.1	2515	10	BC006577	Mus muscu
23	40.6	7.1	281497	2	AC108494	Homo sapi
24	40.6	7.1	311018	2	AC093684	Homo sapi
25	39.2	6.8	182150	9	HSG278N14	Human DNA
26	38.8	6.7	2379	3	DD31	D. discoideu
27	38.8	6.7	30614	9	AB045362	Homo sapi
28	38.8	6.7	146423	2	AC073856	Homo sapi
29	38.8	6.7	180742	9	AL359753	Human DNA
30	38.8	6.7	194529	2	AC036146	Mus muscu
31	38.8	6.7	300000	9	AP002534	Homo sapi
32	38.6	6.7	91825	2	AC019854	Drosophil
33	38.6	6.7	200978	3	AC010060	Drosophil
34	38.6	6.7	274585	3	AE003533	Drosophil
35	38.4	6.7	1134	9	HSM802368	Homo sapi
36	38.4	6.7	105100	9	HS67A5	Human DNA
37	38.4	6.7	125020	9	AF429315	Homo sapi
38	38.4	6.7	293181	1	AP001119	Buchnera
39	38.2	6.6	28586	3	CEC1A46	Caenorhabdi
40	38.2	6.6	34584	8	SPCC1795	S.pombe c
41	38.2	6.6	171478	9	AC016542	Homo sapi
42	38.2	6.6	219728	2	AL645987	Mus muscu
43	38	6.6	108908	3	PFMAL3P8	Plasmodiu
44	38	6.6	130764	2	AC094324	Rattus no
45	37.8	6.6	5173	9	AF426432	Homo sapi

## ALIGNMENTS

RESULT 1

AL390777

LOCUS

DEFINITION

AL390777

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

AL390777 Human DNA sequence from clone RP11-301F14 on chromosome 9, complete sequence.  
182914 bp DNA linear PRI 06-JUN-2001

AL390777

AL390777.13 GI:14329989

HIG.

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 182914)

Brown, J.

Direct Submission

Submitted (02-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire,

CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk

requests: clonerequest@sanger.ac.uk

On Jun 8, 2001 this sequence version replaced qi:14727213.

During sequence assembly data is compared from overlapping clones.

Where differences are found these are annotated as variations

together with a note of the overlapping clone name. Note that the

variation annotation may not be found in the sequence submission





## RESULT

### Quarry Match

Best: Local Similarity 96.28; Score 367.4; DB 4; Length 384;  
Matches 379; Conservative 1.0; Pred. No. 5.1e-99;

[illegible]

Oy 60 agtaccagttaccataaacgaatgcacactgctggcgttcaggccagtgcgcacactggtgcc 50  
bb 1 agcgggagcagaagcttaagccaaagcccagaagagtgccagctgcgcacactggtgcc 59

Db 51 agtaccagttaccacaaataacagtgccagtgccagtgccagaccagtcggttgacttcagtgcct 113

.20 ggtccagcctgaccgcacactctcacatttggcctcttcgtgcaccttggtagagctgg 179  
 .21 ggtaccagcctgaccgcacactctcacatttggcctcttcgtgcaccttggtagagctgg 179  
 Db

[illegible]

1. yccagcaccagtggcagctctggcgccctgtggcttcctcctacaagtggacatttttagatat 240

240 ttttaactctgcaggtcttcttcgaaggatgcatcctcagaacctactcaaca 299  
 |||||||  
 241 ttctaactctgcaggtcttcttcgaaggatgcatcctcagaacctactcaaca 300  
 |||||||  
 300 cagcaccttaggcagccacattcataattgaagtggacactctgcattaaactattt 359  
 |||||||  
 301 cagcaccttaggcagccacataaatcaatgaagtggacactctgcattaratctatt 360  
 |||||||  
 360 gccattaaaaaaaataaaaaa 382  
 |||||||  
 361 gccattcaaaaaaaaaaaaaa 383

### RESULTS

US-09-020-956-82  
: Sequence 82, Application US/09020956  
: Patent No. 6261562  
: GENERAL INFORMATION:  
: APPLICANT: Xu, Jiangchun  
: APPLICANT: Dillin, Davin C.  
: TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE CANCER AND METHOD  
: NUMBER OF SEQUENCES: 178  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: SEED and BERRY LLP  
: STREET: 5300 Columbia Center, 701 Fifth Avenue  
: CITY: Seattle  
: STATE: WA  
: COUNTRY: USA  
: ZIP: 98104  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: Floppy disk  
: COMPUTER: IBM PC compatible  
: OPERATING SYSTEM: PC-DOS/MS-DOS  
: SOFTWARE: Patent In Release #1.0, Version #1.30  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/09-020,956  
: FILING DATE: 09-FEB-1998  
: CLASSIFICATION:  
: ATTORNEY/AGENT INFORMATION:  
: NAME: Makl, David J.  
: REGISTRATION NUMBER: 31,392  
: REFERENCE/DOCKET NUMBER: 210121.427C2  
: TELECOMMUNICATION INFORMATION:  
: TELEPHONE: (206) 622-4900  
: TELEFAX: (206) 682-6031  
: INFORMATION FOR SEQ ID NO: 82:  
: SEQUENCE CHARACTERISTICS:  
: LENGTH: 383 base pairs  
: TYPE: nucleic acid  
: STRANDEDNESS: single  
: TOPOLOGY: linear  
: MOLECULE TYPE: CDNA  
: ORIGINAL SOURCE:  
: ORGANISM: Homo sapiens  
: US-CG-020-956-82

```

Query: Match
Best: Local Similarity 92.9%
Matches: 333 Score 354.8; DB 4; Length 383;
Pred No 250.0

```

[illegible]

59  
50

61 AGTACCAGTACCAATACA - TGCAGTGCCAGTGCCAGCAGCAGCTGCTTT  
bb

120 ggtgccagcctgacccgcactatcacatttgagctcttcgcctggaccttgaatgagact

b  
170 GGTCGAGCCTGACGCCACTCTCACATTTGGGCTCTTCGTGGCCCTTGGTGGAGCTGT 176



Oy 180 gccagaccagtgagcagctctggtgctgtgtttctctcaaaagtgagatttagatat 239  
Db 180 gccagaccagtgagcagctctggtgctgtgtttctctcaaaagtgagatttagatat 239  
Oy 240 tgtaaatccctgccagctcttctctcaagccaggggtgcatctcagaacactactcaaca 299  
Db 240 tgtaaatccctgccagctcttctctcaagccaggggtgcatctcagaacactactcaaca 299  
Oy 300 cagcactctgagcagccactatcaatcaattgaattgacactctgcattaaatctatt 359  
Db 300 cagcactctgagcagccactatcaatcaattgaattgacactctgcattaaatctatt 359  
Oy 360 gccattaaaaaataaaaaaa 382  
Db 360 gccattaaaaaataaaaaaa 382

RESULT 4  
US-09-030-607-82  
: Sequence 82, Application US/09030607  
: Patent No. 6262245  
: GENERAL INFORMATION:  
: APPLICANT: Xu, Jiangchun  
: APPLICANT: Dillon, David C.  
: TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE CANCER AND METHODS FO  
: NUMBER OF SEQUENCES: 224  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: SEED AND BERRY LLP  
: STREET: 6300 Columbia Center, 701 Fifth Avenue  
: CITY: Seattle  
: STATE: WA  
: COUNTRY: USA  
: ZIP: 98104  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: Floppy disk  
: COMPUTER: IBM PC compatible  
: OPERATING SYSTEM: PC-DOS/MS-DOS  
: SOFTWARE: PatentIn Release #1.0, Version #1.30  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/09/030.607  
: FILING DATE: 25-FEB-1998  
: CLASSIFICATION:  
: ATTORNEY/AGENT INFORMATION:  
: NAME: Maki, David J.  
: REGISTRATION NUMBER: 31,392  
: REFERENCE/DOCKET NUMBER: 210121.427C3  
: TELECOMMUNICATION INFORMATION:  
: TELEPHONE: (206) 522-4900  
: TELEFAX: (206) 582-6031  
: INFORMATION FOR SEQ ID NO: 82:  
: SEQUENCE CHARACTERISTICS:  
: LENGTH: 383 base pairs  
: TYPE: nucleic acid  
: STRANDEDNESS: single  
: TOPOLOGY: linear  
: MOLECULE TYPE: cDNA  
: ORIGINAL SOURCE:  
: ORGANISM: Homo sapiens  
: US-09-030-607-82

Query Match 92.9% Score 354.8; DB 4; Length 383;  
Best Local Similarity 98.7% Pred. No. 2.6e-95;  
Matches 378; Conservative 0; Mismatches 3; Indels 2; Gaps 2;

Oy 1 agcggagagagaagctaaagccaaagcccaag-agagtgccagtgccagcactggtgcc 59  
Db 1 agcggagagagaagctaaagccaaagcccaag-agagtgccagtgccagcactggtgcc 60  
Oy 60 agtaccagtagcaccataaagccagtgccagtgccagcaccagtggtggttcagtgct 119  
Db 61 agtaccagtagcaccataaagccagtgccagtgccagcaccagtggtggttcagtgct 119  
Oy 120 ggtgccagcctgaccgcccactctcaacatttgagctcttcgctggccttggtgagcctgct 179  
Db 120 ggtgccagcctgaccgcccactctcaacatttgagctcttcgctggccttggtgagcctgct 179  
Oy 180 gccagcaccagtgagcagctctggtgctgtgtttctctcaaaagtgagatttagatat 239  
Db 180 gccagcaccagtgagcagctctggtgctgtgtttctctcaaaagtgagatttagatat 239

Oy 120 ggtgccagcctgaccgcccactctcaacatttgagctcttcgctggccttggtgagcctgct 179  
Db 120 ggtgccagcctgaccgcccactctcaacatttgagctcttcgctggccttggtgagcctgct 179  
Oy 180 gccagcaccagtgagcagctctggtgctgtgtttctctcaaaagtgagatttagatat 239  
Db 180 gccagcaccagtgagcagctctggtgctgtgtttctctcaaaagtgagatttagatat 239  
Oy 240 tgtaaatccctgccagctcttctctcaagccaggggtgcatctcagaacactactcaaca 299  
Db 240 tgtaaatccctgccagctcttctctcaagccaggggtgcatctcagaacactactcaaca 299  
Oy 300 cagcactctgagcagccactatcaatcaattgaattgacactctgcattaaatctatt 359  
Db 300 cagcactctgagcagccactatcaatcaattgaattgacactctgcattaaatctatt 359  
Oy 360 gccattaaaaaataaaaaaa 382  
Db 360 gccattaaaaaataaaaaaa 382

RESULT 5  
US-09-439-313-82  
: Sequence 82, Application US/09439313  
: Patent No. 6329505  
: GENERAL INFORMATION:  
: APPLICANT: Xu, Jiangchun  
: APPLICANT: Dillon, David C.  
: APPLICANT: Mitcham, Jennifer L.  
: APPLICANT: Harlocker, Susan Louise  
: APPLICANT: Jiang Yuqi  
: APPLICANT: Reed, Steven G.  
: APPLICANT: Kalos, Michael  
: APPLICANT: Fanger, Gary  
: APPLICANT: Retter, Mark  
: APPLICANT: Solk, John  
: APPLICANT: Day, Craig  
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND  
: FILE REFERENCE: 210121.427C9  
: CURRENT APPLICATION NUMBER: US/09/439.313  
: CURRENT FILING DATE: 1999-11-12  
: NUMBER OF SEQ ID NOS: 575  
: SOFTWARE: FastSeq for Windows Version 3.0  
: SEQ ID NO 82  
: LENGTH: 383  
: TYPE: DNA  
: ORGANISM: Homo sapien  
: NAME/KEY: misc\_feature  
: LOCATION: (1)...(383)  
: OTHER INFORMATION: n = A,T,C or G  
: US-09-439-313-82

Query Match 92.9% Score 354.8; DB 4; Length 383;  
Best Local Similarity 98.7% Pred. No. 2.6e-95;  
Matches 378; Conservative 0; Mismatches 3; Indels 2; Gaps 2;

Oy 1 agcggagagagaagctaaagccaaagcccaag-agagtgccagtgccagcactggtgcc 59  
Db 1 agcggagagagaagctaaagccaaagcccaag-agagtgccagtgccagcactggtgcc 60  
Oy 60 agtaccagtagcaccataaagccagtgccagtgccagcaccagtggtggttcagtgct 119  
Db 61 agtaccagtagcaccataaagccagtgccagtgccagcaccagtggtggttcagtgct 119  
Oy 120 ggtgccagcctgaccgcccactctcaacatttgagctcttcgctggccttggtgagcctgct 179  
Db 120 ggtgccagcctgaccgcccactctcaacatttgagctcttcgctggccttggtgagcctgct 179  
Oy 180 gccagcaccagtgagcagctctggtgctgtgtttctctcaaaagtgagatttagatat 239  
Db 180 gccagcaccagtgagcagctctggtgctgtgtttctctcaaaagtgagatttagatat 239

Db 180 gccagcaccagtgagcctgctgctggtttctctctacaagtgagatctttaaata 239  
Oy 240 tggtaactcctgagcagctctctctcaagccaggggtgcatctctcaaaaactactcaaca 299  
Db 240 tggtaactcctgagcagctctctctcaagccaggggtgcatctctcaaaaactactcaaca 299  
Oy 300 cagcactctaggcagcactatcaatcaattgaagttgacactctcattaaatctattt 359  
Db 300 cagcactctaggcagcactatcaatcaattgaagttgacactctcattaaatctattt 359  
Oy 360 gccattaaaaaataaaaaaa 382  
Db 360 gccatttcaaaaaaataaaaa 382

## RESULT 5

US-09-020-956-73  
: Sequence 73, Application US/09020956  
: Patent No. 6261562  
: GENERAL INFORMATION:  
: APPLICANT: Xu, Jiangchun  
: APPLICANT: Dillon, Davin C.  
: TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE CANCER AND METHODS FO  
: NUMBER OF SEQUENCES: 178  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: SEED AND BERRY LLP  
: STREET: 6300 Columbia Center, 701 Fifth Avenue  
: CITY: Seattle  
: STATE: WA  
: COUNTRY: USA  
: ZIP: 98104  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: Floppy disk  
: COMPUTER: IBM PC compatible  
: OPERATING SYSTEM: PC-DOS/MS-DOS  
: SOFTWARE: PatentIn Release #1.0, Version #1.30  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/09/020.956  
: FILING DATE: 09-FEB-1998  
: CLASSIFICATION:  
: ATTORNEY/AGENT INFORMATION:  
: NAME: Maki, David J.  
: REGISTRATION NUMBER: 31,392  
: REFERENCE/DOCKET NUMBER: 210121.427C2  
: TELECOMMUNICATION INFORMATION:  
: TELEPHONE: (206) 622-4900  
: TELEFAX: (206) 682-6031  
: INFORMATION FOR SEQ ID NO: 73:  
: SEQUENCE CHARACTERISTICS:  
: LENGTH: 499 base pairs  
: TYPE: nucleic acid  
: STRANDEDNESS: single  
: TOPOLOGY: linear  
: MOLECULE TYPE: cDNA  
: ORIGINAL SOURCE:  
: ORGANISM: Homo sapiens  
: US-09-020-956-73

Query Match 87.9%; Score 335.6; DB 4; Length 499;  
Best Local Similarity 98.8%; Pred. No. 1.3e-89;  
Matches 338; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
Oy 41 cagtgccagcactggtgccagctaccagttaccataacagtgccagtgcacgtgccagcagcac 100  
Db 1 CAGTGGCCAGCACTGGTGGCCAGTACCAGTACCATAACAGTGGCCAGTGGCCAGTGGCCAGCAC 50  
Oy 101 cagtgagcttcagctgagtgagcagcctgacccgacactctcaatttgagctcttcgc 160  
Db 61 CAGTGGTGGCTTCAGTGGTGGTGGCCAGCTGACCGCCACTCTCACATTTGGGCTCTTCGC 120  
Oy 161 tggccttggtgagagctggtgagcagcaccagtgccagctctgagctgagctcttctccta 220

Db 121 TGGCCTTGGTGGAGCTGGTGGCCAGCAGTGCCAGCTCTGGTGGCTGGTTCCTCTCTA 180  
Oy 221 caagtgaatttgaatatattgtaactctgacagctcttctcttctcaagccaggtgcatc 280  
Db 181 CAAGTGAGATTTAGATATTTGTTAACTCTGCCAGTCTTCTTCTTCAAGCCAGGTGTCATC 240  
Oy 281 ctcaaaaacctactcaaacacagcactctaggcagcactatcaatcaattgaatttaaca 340  
Db 241 CTCAGAAACCTACTCAACACAGCACTCTAGGCAGCCACTATCAATCAATTAAGTTGACA 300  
Oy 341 ctctgcattaaatctatttgccattaaaaaataaaaaaa 382  
Db 301 CTCIGCAITAAATCTATTGGCATTTCGAAAAAATAAAAAA 342

## RESULT 7

US-09-030-607-73  
: Sequence 73, Application US/09030607  
: Patent No. 6262245  
: GENERAL INFORMATION:  
: APPLICANT: Xu, Jiangchun  
: APPLICANT: Dillon, Davin C.  
: TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE CANCER AND METHODS  
: NUMBER OF SEQUENCES: 224  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: SEED AND BERRY LLP  
: STREET: 6300 Columbia Center, 701 Fifth Avenue  
: CITY: Seattle  
: STATE: WA  
: COUNTRY: USA  
: ZIP: 98104  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: Floppy disk  
: COMPUTER: IBM PC compatible  
: OPERATING SYSTEM: PC-DOS/MS-DOS  
: SOFTWARE: PatentIn Release #1.0, Version #1.30  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/09/030.607  
: FILING DATE: 25-FEB-1998  
: CLASSIFICATION:  
: ATTORNEY/AGENT INFORMATION:  
: NAME: Maki, David J.  
: REGISTRATION NUMBER: 31,392  
: REFERENCE/DOCKET NUMBER: 210121.427C3  
: TELECOMMUNICATION INFORMATION:  
: TELEPHONE: (206) 622-4900  
: TELEFAX: (206) 682-6031  
: INFORMATION FOR SEQ ID NO: 73:  
: SEQUENCE CHARACTERISTICS:  
: LENGTH: 499 base pairs  
: TYPE: nucleic acid  
: STRANDEDNESS: single  
: TOPOLOGY: linear  
: MOLECULE TYPE: cDNA  
: ORIGINAL SOURCE:  
: ORGANISM: Homo sapiens  
: US-09-030-607-73

Query Match 87.9%; Score 335.6; DB 4; Length 499;  
Best Local Similarity 98.8%; Pred. No. 1.3e-89;  
Matches 338; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
Oy 41 cagtgccagcactggtgccagctaccagttaccataacagtgccagtgcacgtgccagcagcac 100  
Db 1 CAGTGGCCAGCACTGGTGGCCAGTACCAGTACCATAACAGTGGCCAGTGGCCAGTGGCCAGCAC 60  
Oy 101 cagtgagcttcagctgagtgagcagcctgacccgacactctcaatttgagctcttcgc 160  
Db 61 CAGTGGTGGCTTCAGTGGTGGTGGCCAGCTGACCGCCACTCTCACATTTGGGCTCTTCGC 120  
Oy 161 tggccttggtgagagctggtgagcagcaccagtgccagctctgagctgagctcttctccta 220

[illegible]

PRECEDENT 8

```

US-09-439-313-73
RESOLVING
: Sequence 73, Application US/09439313
: Patent No. 5329505
: GENERAL INFORMATION:
: APPLICANT: Xu, Jiangchun
: APPLICANT: Dillon, Davin C.
: APPLICANT: Mitcham, Jennifer L.
: APPLICANT: Harlocke, Susan Louise
: APPLICANT: Jiang Yuqi
: APPLICANT: Reed, Steven G.
: APPLICANT: Kalos, Michael
: APPLICANT: Fanger, Gary
: APPLICANT: Retter, Mark
: APPLICANT: Solk, John
: APPLICANT: Day, Craig
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE PREPARATION OF A POLYMERIZATION PRODUCT
: FILE REFERENCE: 210121.427C9
: CURRENT APPLICATION NUMBER: US/09/439-313-73
: CURRENT FILING DATE: 1999-11-12
: NUMBER OF SEQ ID NOS: 575
: SOFTWARE: FASTSEQ for Windows Version 1.0
: SEQ ID NO 73
: LENGTH: 499
: TYPE: DNA
: ORGANISM: Homo sapien
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (1)..(499)
: OTHER INFORMATION: n = A,T,C or G
US-09-439-313-73

```

Query Match	87.94:	Score	335.6:	DB 4:	Length	499:
Post Local Similarity	98.8%:	Prod. No.	1.3e-89:			
Matches 338:	Conservative	0:	Mismatches	4:	Indels	0:
Oy	41	cagtgccagcactggtgcaggtaccaggtaccataaacagtgccagtgccagtgccagcacc	100			
Db	1	cagtgccagcactggtgcaggtaccaggtaccataaacagtgccagtgccagtgccagcacc	50			
Oy	101	cagtggtgggtctcagtgctggtgcagagcctgaaccgcaactctccaatcttggtgctcttcgc	160			
Db	61	cagtggtgcttcagtgctggtgcagagcctgaaccgcaactctccaatcttggtgctcttcgc	120			
Oy	161	tggccttggtggagctggtgcgcagcaccagtgctgcagctctgctgctgtgggtttctccta	220			
Db	121	tggccttggtggagctggtgcgcagcaccagtgctgcagctctgctgctgtgggtttctccta	180			
Oy	221	caagtgagattttagatattgtttaactctgcagctctttctcttccaagccagaggtgcacc	280			
Db	181	caagtgagattttagatattgtttaactctgcagctctttctcttccaagccagaggtgcacc	240			
Oy	281	ctcagaacactcactcaacacagcactcttaggcagccactatcaatcaattgaagtgcaca	340			
Db	241	ctcagaacactcactcaacacagcactcttaggcagccactatcaatcaattgaagtgcaca	300			

Qy 341 CTCGcattaaatctatttgcattataaaaaaaaaa 382  
 |||||  
 Db 301 CTCGcattaaatctatttgcatttctataaaaaaaaaa 342

11153d b

```

US-08-718-661-1
: Sequence 1, Application US/08718661
: Patent No. 5876972
: GENERAL INFORMATION:
: APPLICANT:
: TITLE OF INVENTION: Nucleic acid molecules coding for mammalian
:   OF INVENTION: tumor suppressor proteins and methods for their isolation
: NUMBER OF SEQUENCES: 15
: COMPUTER READABLE FORM: disk
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/718,661
: FILING DATE:
: CLASSIFICATION: 530
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2790 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA to mRNA
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 542..2545
US-08-718-661-1

```

Query Match	11.9%;	Score 45.4;	DB 2;	Length 2790;
Best Local Similarity	61.3%;	Pred. No. 0.00071;		
Matches	71.	Conservative	0;	Mismatches 45;
			Indels	0;
			Gaps	0;

Qy 3 gcggagcaaaagctaaagcccaagagagtgcaagtggcagcacttgatgccagt 62  
275 GCACAGCCTGATCTGAGCGAGCGACGAGCGAGCCAGCAGCAGCAGCAGA 2334  
Qy 63 accagtaaccaataacagtcgcagtcgcagtcacacagtcgggtgttcagtcgtcg 121  
2335 cggcggcggcggcggcggcggcggcggcggcggcggcggcggcggcggcggcggc 2393

101 J. H. S. 36

```

1  US-08-317-522A-1
2  : Sequence 1, Application US/08317522A
3  : Patent No. 5599918
4  : GENERAL INFORMATION:
5  : APPLICANT: Fukuda, Michiko N.
6  : TITLE OF INVENTION: Trophinin and Trophinin-Assisting
7  : TITLE OF INVENTION: Proteins
8  : NUMBER OF SEQUENCES: 13
9  : CORRESPONDENCE ADDRESS:
10 : ADDRESSEE: Campbell and Flores
11 : STREET: 4370 La Jolla Village Drive, Suite 700
12 : CITY: San Diego
13 : STATE: California
14 : COUNTRY: USA
15 : ZIP: 92122
16 : COMPUTER READABLE FORM:
17 : MEDIUM TYPE: Floppy disk
18 : COMPUTER: IBM PC compatible
19 : OPERATING SYSTEM: PC-DOS/MS-DOS
20 : SOFTWARE: Patent In Release #1.0, Version #1.25
21 : CURRENT APPLICATION DATA:
22 : APPLICATION NUMBER: US/08/317,522A

```





Job time: 22736 sec

```

DB 931 AACACCACTAGTGGTTTGGTGGTCTATGAGCA 964
      | | | | | | | | | | | | | | | | | |
RESULT 15
US-08-808-599A-1
: Sequence 1, Application US/08808599A
: Patent No. 6111089
: GENERAL INFORMATION:
: APPLICANT: Fukuda, Michiko N.
: TITLE OF INVENTION: Trophinin, Trophinin-Assisting
: TITLE OF INVENTION: Proteins and Methods to Inhibit Implantation
: NUMBER OF SEQUENCES: 41
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Campbell & Flores LLP
: STREET: 4370 La Jolla Village Drive, Suite 700
: CITY: San Diego
: STATE: California
: COUNTRY: USA
: ZIP: 92122
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/808,599A
: FILING DATE: 28-FEB-1997
: CLASSIFICATION: 424
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/317,522
: FILING DATE: 04-OCT-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/439,818
: FILING DATE: 12-MAY-1995
: ATTORNEY/AGENT INFORMATION:
: NAME: Campbell, Cathryn A.
: REGISTRATION NUMBER: 31,815
: REFERENCE/DOCKET NUMBER: P-LA 2256
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (619) 535-9001
: TELEFAX: (619) 535-8949
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2524 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 28..2275
US-08-808-599A-1

Query Match 11.4%; Score 43.6; DB 3; Length 2524;
Best Local Similarity 55.2%; Pred. No. 0.0023;
Matches 85; Conservative 0; Mismatches 69; Indels 0; Gaps 0:

Qy 36 agtggcagtgccagcactggctgagccagtcaccagtaaccataaagtcagtcagtcagtcagtc 95
   || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 811 AGCACCAGTGCCTGGCTTTGGTGGTCTACACTAAGTACCAGCACTGCTTTGATGGCTCTCC 870
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 96 agcaccagtggtggtcagtcgctggtggtggtggtggtggtggtggtggtggtggtggtggtc 155
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 871 AGCACTGGTGGCTTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 189
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 156 ttcgctggccttgggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggt 189
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 931 AACACCACTAGTGGTTTGGTGGTCTATGAGCA 964
```

Search completed: April 29, 2002, 19:55:38

us-09-248-178-59.std.rni

Tue Apr 30 14:17:53 2002

\_\_\_\_\_





REFERENCE	2 (bases 1 to 2064)	
AUTHORS	Gu.J.R.	
TITLE	Direct Submission	
JOURNAL	Submitted (10-MAR-1997) National Laboratory for Oncogenes and Related Genes, Shanghai Cancer Institute, No. 25, Ln. 2200, Xie-Tu Rd., Shanghai 200032, People's Republic of China	
FEATURES	Location/Qualifiers	
Source	1..2064	
gene	/organism="Homo sapiens"	
CDS	/db_xref="taxon:9606"	
	/gene="JCL-1"	
	57..1897	
	/gene="JCL-1"	
	/codon_start=1	
	/product="hepatocellular carcinoma associated protein"	
	/protein_id="AA000728.1"	
	/db_xref="GI:4099969"	
	/translation="MSDISESGAGLRFQAESEKSSMMOTLLITVONVEVETPK ASKALEVEDQAVKSKVSKATEVSKIPAREAPATQASSTIOLITDQVLAENKSL AADTKQNDPQAVTMPATETKKVSHVADTKVNTKAQETEAAPSOAPADEPEPEASAA QSOENQDTRPKVAKKARKVKHLDGEDSSQASGTTGGRVSKALMASMARRAS RGPFAFWRASRTRLAWARALLSLSPKARRGKARRRAKLOSQEPAPPRDV ALLOGRANDLYLLAKDOTKIPKRSMDLKDIIKEVTDVYPIEIERAGYSLEKVFGI QLEIKDNHLYILLSTLEPTDAGILGTIKDSPKGLGLMLVLLSTFIEMNGNRSSEAVIM EVLRKGLRPGIHHSLFGVKKLIITDEFVKQYLDYARVPSNPPEYEFFWGLRSYIE TSKMKVLKFACKVQKDPKWAAYOQAEADLKAAAEAAKAEAKARAEIRAMGILG SENNAAGPCWDEADIGPWAKARIOAGAEAKAEOESGASTGASTINNSASASASTS GGFSAAGSLTATLIFGLFAGLGGAGASTSGSSGACGFSYK"	
BASE COUNT	564 a 523 c 570 g 407 t	
ORIGIN		
Query Match	99.2%; Score 378.8; DB 9; Length 2064;	
Best Local Similarity	99.5%; Pred. No. 3.7e-99;	
Matches	380; Conservative 0; Mismatches 2; Indels 0; Gaps 0;	
Oy	1 agcgaggagcagaagctaaagccaaagcgaagagtgccagtcgagcagcactgggtgccca 60	
Db	1661 AGCGGGAGCAGAGAGCTAAAGCCAAAGCCAGAGTGGCAGCTGGCGACCTGGTGCCA 1720	
Oy	51 gtaccagttaccataaagcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtc 120	
Db	1721 GTACCAGTACCATAAAGCAGTGGCAGTGGCAGTGGCAGTGGCAGTGGCAGTGGCAGT 1780	
Oy	121 gtccagcctgacgcagcactctcacatttgggctcttcgcttggccttggcagcagtcgt 180	
Db	1781 GTGCCAGCCTGACGCCACTCTCACATTGGGCTCTTCGCTGGCCTTGGTGGAGCTGGTG 1840	
Oy	181 ccagcagcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtc 240	
Db	1841 CCAGCAGCAGTGGCAGCTCTGGTGGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1900	
Oy	241 gttaatcctgacgcagtcctttctcttcaagccaggggtgagtcctcctcagaacactactcaacac 300	
Db	1901 GTTAATCTGTCAGTCTTCTCTTCAAGCCAGGGTGGATCTCTCAGAACCTTACTCAACAC 1960	
Oy	301 agcaactctagcagcagcactatcaatcaatgaagttagcactctgcattaaactatttg 360	
Db	1961 AGCACTCTAGGCAGCCACTATCAATCAATTCAGTTGACACTCTGCATTAAATCTATTG 2020	
Oy	361 ccattaaaaaataaaaaa 382	
Db	2021 CCATTTCATAAAAAAATAAAAAA 2042	
RESULT	3	
LOCUS	BC000304	
DEFINITION	Homo sapiens, hepatocellular carcinoma associated protein; breast cancer associated gene 1, clone MGC:8386 IMAGE:2820442, mRNA.	
ACCESSION	BC000304	
KEYWORDS	complete cds.	
SOURCE	human.	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
AUTHORS	1 (bases 1 to 2064)	
TITLE	Jin,C.L., Wang,D.Y., Wan,D.F. and Gu,J.R.	
JOURNAL	Hepatocellular Carcinoma Associated Gene JCL-1	
published		
organism="Homo sapiens"		
/db_xref="taxon:9606"		
/clone="Xp11B6"		
/clone_lib="placenta"		
13..1833		
CDS		
/codon_start=1		
/product="hypothetical protein 11B6"		
/protein_id="CAC12410.1"		
/db_xref="GI:11967746"		
/translation="MSDISESGAGLRFQAESEKSSMMOTLLITVONVEVETPK ASKALEVEDQAVKSKVSKATEVSKIPAREAPATQASSTIOLITDQVLAENKSL AADTKQNDPQAVTMPATETKKVSHVADTKVNTKAQETEAAPSOAPADEPEPEASAA QSOENQDTRPKVAKKARKVKHLDGEDSSQASGTTGGRVSKALMASMARRAS RGPFAFWRASRTRLAWARALLSLSPKARRGKARRRAKLOSQEPAPPRDV ALLOGRANDLYLLAKDOTKIPKRSMDLKDIIKEVTDVYPIEIERAGYSLEKVFGI QLEIKDNHLYILLSTLEPTDAGILGTIKDSPKGLGLMLVLLSTFIEMNGNRSSEAVIM EVLRKGLRPGIHHSLFGVKKLIITDEFVKQYLDYARVPSNPPEYEFFWGLRSYIE TSKMKVLKFACKVQKDPKWAAYOQAEADLKAAAEAAKAEAKARAEIRAMGILG SENNAAGPCWDEADIGPWAKARIOAGAEAKAEOESGASTGASTINNSASASASTS GGFSAAGSLTATLIFGLFAGLGGAGASTSGSSGACGFSYK"		
BASE COUNT	528 a 507 c 553 g 398 t	
ORIGIN		
Query Match	99.5%; Score 380; DB 9; Length 1986;	
Best Local Similarity	100.0%; Pred. No. 1.7e-99;	
Matches	380; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
Oy	1 agcgaggagcagaagctaaagccaaagcgaagagtgccagtcgagcagcactgggtgccca 60	
Db	1607 AGCGGGAGCAGAGAGCTAAAGCCAAAGCCAGAGAGTGGCAGCTGGCGACCTGGTGCCA 1656	
Oy	61 gtaccagttaccataaagcagtcgagtcgagtcgagtcgagtcgagtcgagtcgagtcg 120	
Db	1667 GTACCAGTACCATAAAGCAGTGGCAGTGGCAGTGGCAGTGGCAGTGGCAGTGGCAGT 1726	
Oy	121 gtccagcctgacgcagcactctcacatttgggctcttcgcttggccttggcagcagtcgt 180	
Db	1727 GTGCCAGCCTGACGCCACTCTCACATTGGGCTCTTCGCTGGCCTTGGTGGAGCTGGTG 1785	
Oy	181 ccagcagcagtcgagcagtccttggccttggccttggccttggccttggccttggccttgg 240	
Db	1787 CCAGCAGCAGTGGCAGCTCTGGTGGCTGGTGGCTGGTGGCTGGTGGCTGGTGGCTGG 1845	
Oy	241 gttaatcctgacgcagtccttcttcaagccagcagtcagtcagtcagtcagtcagtcagtc 300	
Db	1847 GTTAATCTGTCAGTCTTCTCTTCAAGCCAGGGTGGATCTCTCAGAACCTTACTCAACAC 1906	
Oy	301 agcaactctagcagcagcactatcaatcaatgaagttagcactctgcattaaactatttg 360	
Db	1907 AGCACTCTAGGCAGCCACTATCAATCAATTCAGTTGACACTCTGCATTAAATCTATTG 1966	
Oy	361 ccattaaaaaataaaaaa 380	
Db	1967 CCATTTCATAAAAAAATAAAAAA 1986	
RESULT	2	
LOCUS	HSU92544	
DEFINITION	Human hepatocellular carcinoma associated protein (JCL-1) mRNA, complete cds.	
ACCESSION	U92544	
VERSION	U92544.1	
KEYWORDS	GI:4099968	
SOURCE	human.	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
AUTHORS	1 (bases 1 to 2064)	
TITLE	Jin,C.L., Wang,D.Y., Wan,D.F. and Gu,J.R.	
JOURNAL	Hepatocellular Carcinoma Associated Gene JCL-1	
published		
organism="Homo sapiens"		
/db_xref="taxon:9606"		
/clone="Xp11B6"		
/clone_lib="placenta"		
13..1833		
CDS		
/codon_start=1		
/product="hypothetical protein 11B6"		
/protein_id="CAC12410.1"		
/db_xref="GI:11967746"		
/translation="MSDISESGAGLRFQAESEKSSMMOTLLITVONVEVETPK ASKALEVEDQAVKSKVSKATEVSKIPAREAPATQASSTIOLITDQVLAENKSL AADTKQNDPQAVTMPATETKKVSHVADTKVNTKAQETEAAPSOAPADEPEPEASAA QSOENQDTRPKVAKKARKVKHLDGEDSSQASGTTGGRVSKALMASMARRAS RGPFAFWRASRTRLAWARALLSLSPKARRGKARRRAKLOSQEPAPPRDV ALLOGRANDLYLLAKDOTKIPKRSMDLKDIIKEVTDVYPIEIERAGYSLEKVFGI QLEIKDNHLYILLSTLEPTDAGILGTIKDSPKGLGLMLVLLSTFIEMNGNRSSEAVIM EVLRKGLRPGIHHSLFGVKKLIITDEFVKQYLDYARVPSNPPEYEFFWGLRSYIE TSKMKVLKFACKVQKDPKWAAYOQAEADLKAAAEAAKAEAKARAEIRAMGILG SENNAAGPCWDEADIGPWAKARIOAGAEAKAEOESGASTGASTINNSASASASTS GGFSAAGSLTATLIFGLFAGLGGAGASTSGSSGACGFSYK"		
BASE COUNT	528 a 507 c 553 g 398 t	
ORIGIN		
Query Match	99.5%; Score 380; DB 9; Length 1986;	
Best Local Similarity	100.0%; Pred. No. 1.7e-99;	
Matches	380; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
Oy	1 agcgaggagcagaagctaaagccaaagcgaagagtgccagtcgagcagcactgggtgccca 60	
Db	1607 AGCGGGAGCAGAGAGCTAAAGCCAAAGCCAGAGAGTGGCAGTGGCAGCTGGTGCCA 1656	
Oy	61 gtaccagttaccataaagcagtcgagtcgagtcgagtcgagtcgagtcgagtcgagtcg 120	
Db	1667 GTACCAGTACCATAAAGCAGTGGCAGTGGCAGTGGCAGTGGCAGTGGCAGTGGCAGT 1726	
Oy	121 gtccagcctgacgcagcactctcacatttgggctcttcgcttggccttggcagcagtcgt 180	
Db	1727 GTGCCAGCCTGACGCCACTCTCACATTGGGCTCTTCGCTGGCCTTGGTGGAGCTGGTG 1785	
Oy	181 ccagcagcagtcgagcagtccttggccttggccttggccttggccttggccttggccttgg 240	
Db	1787 CCAGCAGCAGTGGCAGCTCTGGTGGCTGGTGGCTGGTGGCTGGTGGCTGGTGGCTGG 1845	
Oy	241 gttaatcctgacgcagtccttcttcaagccagcagtcagtcagtcagtcagtcagtcagtc 300	
Db	1847 GTTAATCTGTCAGTCTTCTCTTCAAGCCAGGGTGGATCTCTCAGAACCTTACTCAACAC 1906	
Oy	301 agcaactctagcagcagcactatcaatcaatgaagttagcactctgcattaaactatttg 360	
Db	1907 AGCACTCTAGGCAGCCACTATCAATCAATTCAGTTGACACTCTGCATTAAATCTATTG 1966	
Oy	361 ccattaaaaaataaaaaa 380	
Db	1967 CCATTTCATAAAAAAATAAAAAA 1986	
RESULT	2	
LOCUS	HSU92544	
DEFINITION	Human hepatocellular carcinoma associated protein (JCL-1) mRNA, complete cds.	
ACCESSION	U92544	
VERSION	U92544.1	
KEYWORDS	GI:4099968	
SOURCE	human.	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
AUTHORS	1 (bases 1 to 2064)	
TITLE	Jin,C.L., Wang,D.Y., Wan,D.F. and Gu,J.R.	
JOURNAL	Hepatocellular Carcinoma Associated Gene JCL-1	
published		
organism="Homo sapiens"		
/db_xref="taxon:9606"		
/clone="Xp11B6"		
/clone_lib="placenta"		
13..1833		
CDS		
/codon_start=1		
/product="hypothetical protein 11B6"		
/protein_id="CAC12410.1"		
/db_xref="GI:11967746"		
/translation="MSDISESGAGLRFQAESEKSSMMOTLLITVONVEVETPK ASKALEVEDQAVKSKVSKATEVSKIPAREAPATQASSTIOLITDQVLAENKSL AADTKQNDPQAVTMPATETKKVSHVADTKVNTKAQETEAAPSOAPADEPEPEASAA QSOENQDTRPKVAKKARKVKHLDGEDSSQASGTTGGRVSKALMASMARRAS RGPFAFWRASRTRLAWARALLSLSPKARRGKARRRAKLOSQEPAPPRDV ALLOGRANDLYLLAKDOTKIPKRSMDLKDIIKEVTDVYPIEIERAGYSLEKVFGI QLEIKDNHLYILLSTLEPTDAGILGTIKDSPKGLGLMLVLLSTFIEMNGNRSSEAVIM EVLRKGLRPGIHHSLFGVKKLIITDEFVKQYLDYARVPSNPPEYEFFWGLRSYIE TSKMKVLKFACKVQKDPKWAAYOQAEADLKAAAEAAKAEAKARAEIRAMGILG SENNAAGPCWDEADIGPWAKARIOAGAEAKAEOESGASTGASTINNSASASASTS GGFSAAGSLTATLIFGLFAGLGGAGASTSGSSGACGFSYK"		
BASE COUNT	528 a 507 c 553 g 398 t	
ORIGIN		
Query Match	99.5%; Score 380; DB 9; Length 1986;	
Best Local Similarity	100.0%; Pred. No. 1.7e-99;	
Matches	380; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
Oy	1 agcgaggagcagaagctaaagccaaagcgaagagtgccagtcgagcagcactgggtgccca 60	
Db	1607 AGCGGGAGCAGAGAGCTAAAGCCAAAGCCAGAGAGTGGCAGTGGCAGCTGGTGCCA 1656	
Oy	61 gtaccagttaccataaagcagtcgagtcgagtcgagtcgagtcgagtcgagtcgagtcg 120	
Db	1667 GTACCAGTACCATAAAGCAGTGGCAGTGGCAGTGGCAGTGGCAGTGGCAGTGGCAGT 1726	
Oy	121 gtccagcctgacgcagcactctcacatttgggctcttcgcttggccttggcagcagtcgt 180	
Db	1727 GTGCCAGCCTGACGCCACTCTCACATTGGGCTCTTCGCTGGCCTTGGTGGAGCTGGTG 1785	
Oy	181 ccagcagcagtcgagcagtccttggccttggccttggccttggccttggccttggccttgg 240	
Db	1787 CCAGCAGCAGTGGCAGCTCTGGTGGCTGGTGGCTGGTGGCTGGTGGCTGGTGGCTGG 1845	
Oy	241 gttaatcctgacgcagtccttcttcaagccagcagtcagtcagtcagtcagtcagtcagtc 300	
Db	1847 GTTAATCTGTCAGTCTTCTCTTCAAGCCAGGGTGGATCTCTCAGAACCTTACTCAACAC 1906	
Oy	301 agcaactctagcagcagcactatcaatcaatgaagttagcactctgcattaaactatttg 360	
Db	1907 AGCACTCTAGGCAGCCACTATCAATCAATTCAGTTGACACTCTGCATTAAATCTATTG 1966	
Oy	361 ccattaaaaaataaaaaa 380	
Db	1967 CCATTTCATAAAAAAATAAAAAA 1986	
RESULT	2	
LOCUS	HSU92544	
DEFINITION	Human hepatocellular carcinoma associated protein (JCL-1) mRNA, complete cds.	
ACCESSION	U92544	
VERSION	U92544.1	
KEYWORDS	GI:4099968	
SOURCE	human.	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
AUTHORS	1 (bases 1 to 2064)	
TITLE	Jin,C.L., Wang,D.Y., Wan,D.F. and Gu,J.R.	
JOURNAL	Hepatocellular Carcinoma Associated Gene JCL-1	
published		
organism="Homo sapiens"		
/db_xref="taxon:9606"		
/clone="Xp11B6"		
/clone_lib="placenta"		
13..1833		
CDS		
/codon_start=1		
/product="hypothetical protein 11B6"		
/protein_id="CAC12410.1"		
/db_xref="GI:11967746"		
/translation="MSDISESGAGLRFQAESEKSSMMOTLLITVONVEVETPK ASKALEVEDQAVKSKVSKATEVSKIPAREAPATQASSTIOLITDQVLAENKSL AADTKQNDPQAVTMPATETKKVSHVADTKVNTKAQETEAAPSOAPADEPEPEASAA QSOENQDTRPKVAKKARKVKHLDGEDSSQASGTTGGRVSKALMASMARRAS RGPFAFWRASRTRLAWARALLSLSPKARRGKARRRAKLOSQEPAPPRDV ALLOGRANDLYLLAKDOTKIPKRSMDLKDIIKEVTDVYPIEIERAGYSLEKVFGI QLEIKDNHLYILLSTLEPTDAGILGTIKDSPKGLGLMLVLLSTFIEMNGNRSSEAVIM EVLRKGLRPGIHHSLFGVKKLIITDEFVKQYLDYARVPSNPPEYEFFWGLRSYIE TSKMKVLKFACKVQKDPKWAAYOQAEADLKAAAEAAKAEAKARAEIRAMGILG SENNAAGPCWDEADIGPWAKARIOAGAEAKAEOESGASTGASTINNSASASASTS GGFSAAGSLTATLIFGLFAGLGGAGASTSGSSGACGFSYK"		
BASE COUNT	528 a 507 c 553 g 398 t	
ORIGIN		
Query Match	99.5%; Score 380; DB 9; Length 1986;	
Best Local Similarity	100.0%; Pred. No. 1.7e-99;	
Matches	380; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
Oy	1 agcgaggagcagaagctaaagccaaagcgaagagtgccagtcgagcagcactgggtgccca 60	
Db	1607 AGCGGGAGCAGAGAGCTAAAGCCAAAGCCAGAGAGTGGCAGTGGCAGCTGGTGCCA 1656	
Oy	61 gtaccagttaccataaagcagtcgagtcgagtcgagtcgagtcgagtcgagtcgagtcg 120	
Db	1667 GTACCAGTACCATAAAGCAGTGGCAGTGGCAGTGGCAGTGGCAGTGGCAGTGGCAGT 1726	
Oy	121 gtccagcctgacgcagcactctcacatttgggctcttcgcttggccttggcagcagtcgt 180	
Db	1727 GTGCCAGCCTGACGCCACTCTCACATTGGGCTCTTCGCTGGCCTTGGTGGAGCTGGTG 1785	
Oy	181 ccagcagcagtcgagcagtccttggccttggccttggccttggccttggccttggccttgg 240	
Db	1787 CCAGCAGCAGTGGCAGCTCTGGTGGCTGGTGGCTGGTGGCTGGTGGCTGGTGGCTGG 1845	
Oy	241 gttaatcctgacgcagtccttcttcaagccagcagtcagtcagtcagtcagtcagtcagtc 300	
Db	1847 GTTAATCTGTCAGTCTTCTCTTCAAGCCAGGGTGGATCTCTCAGAACCTTACTCAACAC 1906	
Oy	301 agcaactctagcagcagcactatcaatcaatgaagttagcactctgcattaaactatttg 360	
Db	1907 AGCACTCTAGGCAGCCACTATCAATCAATTCAGTTGACACTCTGCATTAAATCTATTG 1966	
Oy	361 ccattaaaaaataaaaaa 380	
Db	1967 CCATTTCATAAAAAAATAAAAAA 1986	
RESULT	2	
LOCUS	HSU92544	
DEFINITION	Human hepatocellular carcinoma associated protein (JCL-1) mRNA, complete cds.	
ACCESSION	U92544	
VERSION	U92544.1	
KEYWORDS	GI:4099968	
SOURCE	human.	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
AUTHORS	1 (bases 1 to 2064)	
TITLE	Jin,C.L., Wang,D.Y., Wan,D.F. and Gu,J.R.	
JOURNAL	Hepatocellular Carcinoma Associated Gene JCL-1	
published		
organism="Homo sapiens"		
/db_xref="taxon:9606"		
/clone="Xp11B6"		
/clone_lib="placenta"		
13..1833		
CDS		
/codon_start=1		
/product="hypothetical protein 11B6"		
/protein_id="CAC12410.1"		
/db_xref="GI:11967746"		
/translation="MSDISESGAGLRFQAESEKSSMMOTLLITVONVEVETPK ASKALEVEDQAVKSKVSKATEVSKIPAREAPATQASSTIOLITDQVLAENKSL AADTKQNDPQAVTMPATETKKVSHVADTKVNTKAQETEAAPSOAPADEPEPEASAA QSOENQDTRPKVAKKARKVKHLDGEDSSQASGTTGGRVSKALMASMARRAS RGPFAFWRASRTRLAWARALLSLSPKARRGKARRRAKLOSQEPAPPRDV ALLOGRANDLYLLAKDOTKIPKRSMDLKDIIKEVTDVYPIEIERAGYSLEKVFGI QLEIKDNHLYILLSTLEPTDAGILGTIKDSPKGLGLMLVLLSTFIEMNGNRSSEAVIM EVLRKGLRPGIHHSLFGVKKLIITDEFVKQYLDYARVPSNPPEYEFFWGLRSYIE TSKMKVLKFACKVQKDPKWAAYOQAEADLKAAAEAAKAEAKARAEIRAMGILG SENNAAGPCWDEADIGPWAKARIOAGAEAKAEOESGASTGASTINNSASASASTS GGFSAAGSLTATLIFGLFAGLGGAGASTSGSSGACGFSYK"		
BASE COUNT	528 a 507 c 553 g 398 t	
ORIGIN		
Query Match	99.5%; Score 380; DB 9; Length 1986;	
Best Local Similarity	100.0%; Pred. No. 1.7e-99;	
Matches	380; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
Oy	1 agcgaggagcagaagctaaagccaaagcgaagagtgccagtcgagcagcactgggtgccca 60	
Db	1607 AGCGGGAGCAGAGAGCTAAAGCCAAAGCCAGAGAGTGGCAGTGGCAGCTGGTGCCA 1656	
Oy	61 gtaccagttaccataaagcagtcgagtcgagtcgagtcgagtcgagtcgagtcgagtcg 120	
Db	1667 GTACCAGTACCATAAAGCAGTGGCAGTGGCAGTGGCAGTGGCAGTGGCAGTGGCAGT 1726	
Oy	121 gtccagcctgacgcagcactctcacatttgggctcttcgcttggccttggcagcagtcgt 180	
Db	1727 GTGCCAGCCTGACGCCACTCTCACATTGGGCTCTTCGCTGGCCTTGGTGGAGCTGGTG 1785	
Oy	181 ccagcagcagtcgagcagtccttggccttggccttggccttggccttggccttggccttgg 240	
Db	1787 CCAGCAGCAGTGGCAGCTCTGGTGGCTGGTGGCTGGTGGCTGGTGGCTGGTGGCTGG 1845	
Oy	241 gttaatcctgacgcagtccttcttcaagccagcagtcagtcagtcagtcagtcagtcagtc 300	
Db	1847 GTTAATCTGTCAGTCTTCTCTTCAAGCCAGGGTGGATCTCTCAGAACCTTACTCAACAC 1906	
Oy	301 agcaactctagcagcagcactatcaatcaatgaagttagcactctgcattaaactatttg 360	
Db	1907 AGCACTCTAGGCAGCCACTATCAATCAATTCAGTTGACACTCTGCATTAAATCTATTG 1966	
Oy	361 ccattaaaaaataaaaaa 380	
Db	1967 CCATTTCATAAAAAAATAAAAAA 1986	
RESULT	2	
LOCUS	HSU92544	
DEFINITION	Human hepatocellular carcinoma associated protein (JCL-1) mRNA, complete cds.	
ACCESSION	U92544	
VERSION	U92544.1	
KEYWORDS	GI:4099968	
SOURCE	human.	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
AUTHORS	1 (bases 1 to 2064)	
TITLE	Jin,C.L., Wang,D.Y., Wan,D.F. and Gu,J.R.	
JOURNAL	Hepatocellular Carcinoma Associated Gene JCL-1	
published		
organism="Homo sapiens"		
/db_xref="taxon:9606"		
/clone="Xp11B6"		
/clone_lib="placenta"		
13..1833		
CDS		
/codon_start=1		
/product="hypothetical protein 11B6"		
/protein_id="CAC12410.1"		
/db_xref="GI:11967746"		
/translation="MSDISESGAGLRFQAESEKSSMMOTLLITVONVEVETPK ASKALEVEDQAVKSKVSKATEVSKIPAREAPATQASSTIOLITDQVLAENKSL AADTKQNDPQAVTMPATETKKVSHVADTKVNTKAQETEAAPSOAPADEPEPEASAA QSOENQDTRPKVAKKARKVKHLDGEDSSQASGTTGGRVSKALMASMARRAS RGPFAFWRASRTRLAWARALLSLSPKARRGKARRRAKLOSQEPAPPRDV ALLOGRANDLYLLAKDOTKIPKRSMDLKDIIKEVTDVYPIEIERAGYSLEKVFGI QLEIKDNHLYILLSTLEPTDAGILGTIKDSPKGLGLMLVLLSTFIEMNGNRSSEAVIM EV		

```

VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
PEMARK
COMMENT

BC000304.1 GI:12653076
MGC.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 2077)
Straussberg, R.
Direct Submission
Submitted (15-NOV-2000) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: DCTD/DTP
cDNA Library Preparation: Rubin Laboratory
cDNA Library Prepared by: The I.M.A.G.E. Consortium (ILLUM)
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC),
Gaithersburg, Maryland;
Web site: http://www.nisc.nih.gov/
Contact: nisc_mgc@nri.nih.gov
Shevchenko, Y., Wetherby, K.D., Beckstrom-Sternberg, S.M.,
Benjamin, B., Blakesley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S.,
Dietrich, N.L., Guan, X., Gupta, J., Ho, S.-L., Karlins, E., Ledaspi, R.,
Lim, M., Maduro, O.L., Masello, C., Mastrian, S.D., McCloskey, J.C.,
McDowell, J., Pearson, R., Snyder, B., Slantripop, S., Thomas, P.J.,
Tiongson, E.E., Touchman, J.W., Tsurgeon, C., Vogt, J.L., Walker, M.A.,
Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/ILLUM at: http://image.llnl.gov
Series: IRAL Plate: 1 Row: c Column: 17
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 3288022.
Location/Qualifiers
1..2077
/organism="Homo sapiens"
/db_xref="locus:10916"
/db_xref="taxon:9606"
/clone="MGC:8386 IMAGE:2820442"
/tissue_type="Lung, small cell carcinoma"
/clone_lib="NIH_MGC_7"
/lab_host="DH10B-R"
/notes="Vector: pOTB7"
97..1917
/codon_start=1
/product="hepatocellular carcinoma associated protein:
breast cancer associated gene 1"
/protein_id="AAH00304.1"
/db_xref="GI:12653077"
/translation="MSDTSESGAGLTFQAEASEKDSMSMOTLLTVTONVVEPIPK
ASKALVEQDVKSKAGVSKATVEVSKTPAEAPATQASSTIOTDVOVAENKSL
AADTKKONADPOVATPATETKRVSHVADTKVHTKAQETEAAPSOAPADEPEASAA
QSENOQDTPKVKAKKRVKHLDEGSDSDSOASGTTGGRRVSKALMASMPBPAS
RGPIAWRASRTRLAARWALLSLSPKARBGKARBPAAKLOSSOEPEAPPDVR
ALLOGRANRLYLLAKDQTKPIKRSMLKDIKEYIDVPEIIFRAGYSLEKVEGI
QLETKDNHLYIHLSLFDGKLLITDEFVKQYLDYARVPNSPPTFEFFWGLRSYE
EVLRKGLRPIHHSFGVKKLITDEFVKQYLDYARVPNSPPTFEFFWGLRSYE
TSKMKVLKFPACKVQKPKWEAAQYREAMEADLKAASAAEAAKARAEIIFARMGIGLG
SENAAGPCNDEADIGPWAKRIQAGAEAKAKAESGASIGASTINNSASASASTS
GGTFAGASLTLTIFGLFAGLGGAGASTSGSSGACGFSYK"
BASE COUNT 551 a 527 c 587 g 412 t
ORIGIN

Query Match 99.2%; Score 378.8; DB 9; Length 2077;
Best Local Similarity 99.5%; Pred. No. 3.7e-99;
Matches 380; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

1 agcggagcagaactaaagccaaagccaaagagagtgccagtgccagcactggtacca 60
|||||
Db 1691 AGCGGGAGCAGCAAGCTAAAGCCAAAGCCAAAGAGAGTGGCAGTCCAGCACTGGTGCCA 1750
|||||
Oy 61 gtaccagtaccataaacaagtcgaagtcgaagtcgaagtcgaagtcgaagtcgaagtcg 120
|||||
Db 1751 GTACCAAGTACCAATAACAGTGCAGTGCCAGTGCAGCAGCAGCAGTGCAGTGCAGTGC 1810
|||||
Oy 121 gtccagcctgaccgcctctcaatttgggctcttgcctgcttgcctgcttgcctgctg 180
|||||
Db 1811 GTGCCAGCCTGACCGCCACTCTCACATTTGGGCTCTTCGCTGGCCTTGGTGAGCTGGTG 1870
|||||
Oy 181 ccagcagcagtcgagcagtcgagtcgagtcgagtcgagtcgagtcgagtcgagtcg 240
|||||
Db 1871 CCAGCAGCAGTGGCAGCTCTGGTGCCCTGGTGGTTCCTCTACCAAGTGAGATTTAGATAT 1930
|||||
Oy 241 gtttaattccctaccagtccttctcttcaagccaggggtgcatctcacaagaacctactcaaac 300
|||||
Db 1931 GTTAATCTGCCAGTCTTCTCTTCAAGCCAGGGTGCATCTCTCAGAAACCTACTCAACAC 1990
|||||
Oy 301 agcactctagcagcagcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtc 360
|||||
Db 1991 AGCACTCTAGGCAGCCACTATCAATCAATTAATGAAGTTGACACTCTGCATTAATCTATTG 2050
|||||
Oy 361 ccattataaataaataaataaataaataaataaataaataaataaataaataaataaataa 382
|||||
Db 2051 CCATTTCAGAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 2072
|||||

RESULT 4
AX015361 AX015361 1462 bp DNA linear PAT 07-SEP-2000
LOCUS Sequence 20 from Patent WO99511727.
ACCESSION AX015361
VERSION AX015361.1 GI:10041341
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1462)
AUTHORS Schmitt, A., Specht, T., Dahl, E., Hinzmann, B., Rosenthal, A. and
Pillarsky, C.
TITLE Human nucleic acid sequences of normal ovary tissue
JOURNAL Patent: WO 9951727-A 20 14-OCT-1999;
SPECHT ARMIN (DE); SPECHT THOMAS (DE); DAHL EDGAR (DE); HINZMANN
BERND (DE); ROSENTHAL ANDRE (DE); METAGEN GES FUER GENOMFORSCHUN
(DE); PILLARSKY CHRISTIAN (DE)
FEATURES
source
1..1462
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 377 a 343 c 399 g 343 t
ORIGIN

Query Match 98.3%; Score 375.6; DB 6; Length 1462;
Best Local Similarity 99.0%; Pred. No. 3e-98;
Matches 378; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 1 agcgggagcagaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 60
|||||
Db 1055 AGCGGGAGCAGCAAGCTAAAGCCAAAGCCAAAGAGAGTGGCAGTCCAGCACTGGTGCCA 1114
|||||
Oy 61 gtaccagtaccataaacaagtcgaagtcgaagtcgaagtcgaagtcgaagtcgaagtcg 120
|||||
Db 1115 GTACCAAGTACCAATAACAGTGCAGTGCCAGTGCCAGCAGCAGTGGTGGTGGTG 1174
|||||
Oy 121 gtccagcctgaccgcctctcaatttgggctcttgcctgcttgcctgcttgcctgctg 180
|||||
Db 1175 GTGCCAGCCTGACCGCCACTCTCACATTTGGGCTCTTCGCTGGCCTTGGTGAGCTGGTG 1234
|||||
Oy 181 ccagcagcagtcgagcagtcgagtcgagtcgagtcgagtcgagtcgagtcgagtcgagtcg 240
|||||
1 agcgggagcagaactaaagccaaagccaaagagagtgccagtgccagcactggtacca 60
|||||

```

```
Db 1235 CCAGCACCAGTGGCAGCTCTGGTGGCTTCTCCACAGTGAGATTTTATGATAT 1294
Oy 241 gtaatacttcgcagctcttcttcttaagccagggtgcatcctcagaacacctactcaaac 300
Db 1295 GTTAATCTGCGCAGCTCTTCTTCAAGCCAGGGTGCATCTCAGAACCTTACTTCAACAC 1354
Oy 301 agcactctaggcagccactatcaatcaattgaagttagacactctgcattaaatctatttg 350
Db 1355 AGCAGCTTAGGAGGAGCAGCTATCAATCAATTGAGTTGACACTCTGCAITAAATCTATTG 1414
Oy 361 ccattataaaaaaa 382
Db 1415 CCAATTTTCAGAAAAA 1436

RESULT 5
AX337185/c
LOCUS AX337185 446 bp DNA linear PAT 09-JAN-2002
DEFINITION Sequence 7694 from Patent WO0194629.
ACCESSION AX337185
VERSION AX337185.1 GI:18127904
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (sites)
AUTHORS Young,P.E., Augustus,M., Carter,K.C., Ebner,R., Endress,G.,
Horrigan,S., Soppet,D.R. and Weaver,Z.
TITLE Cancer gene determination and therapeutic screening using signature
gene sets
JOURNAL Patent: WO 0194629-A 7694 13-DEC-2001;
Avalon Pharmaceuticals (US)
FEATURES
source
1. .446
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 102 a 111 c 123 g 110 t
ORIGIN

Query Match 96.5% Score 369.2; DB 6; Length 446;
Best Local Similarity 97.9%; Pred. No. 1.9e-96;
Matches 374; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Oy 1 aggcggagcagaagctaaagccaaagcccaagagtgagcagtgccacactggcgcca 60
Db 383 AGCGGGAGCAGAGAGCTAAAGCCAAAGCCCAAGAGAGTGCGCAGTGCCAGCCTGGTGCCA 324
Oy 61 gtaccagtaccataaaccagtgccagtgccagtgccagcaccagtggtggttcagtgctg 120
Db 323 GTACCAGTACCATAACAGTGCAGTGCCAGTGCCAGTCCAGCACCAGTGGTGGCTTCAGTGTG 264
Oy 121 gtgcagcctgacccgcaactctacattggggtcttcgtgcttcgtgcttggtggagcagtg 180
Db 263 GTGCCAGCCTGACCGCCACTCTACATTTGGGCTCTTCGCTGGGCTTGGTGAGCTGGTG 204
Oy 181 ccagcaccagtgagcagctctgagctctgagctctgagctctcctcacaagtgacatttagatatt 240
Db 203 CCAGCACCAGTGGCAGCTCTGGTGGCTTGTGTTTCTTCTTCAAGTGTGAGATTTTATGATAT 144
Oy 241 gtaatacttcgcagctcttcttcttaagccagggtgcatcctcagaacacctactcaaac 300
Db 143 GTTAATCTGCGCAGCTCTTCTTCAAGCCAGGGTGCATCTCAGAACCTTACTTCAACAC 84
Oy 301 agcactctaggcagccactatcaatcaattgaagttagacactctgcattaaatctatttg 350
Db 83 AGCAGCTTAGGAGGAGCAGCTATCAATCAATTGAGTTGACACTCTGCAITAAATCTATTG 24
Oy 361 ccattataaaaaaa 382
Db 23 CCAATTTTCAGAAAAA 2
```

```
RESULT 6
AX106402
LOCUS AX106402 384 bp DNA linear PAT 30-APR-2001
DEFINITION Sequence 183 from Patent WO0125272.
ACCESSION AX106402
VERSION AX106402.1 GI:13922081
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 384)
AUTHORS Xu,J., Skeiky,Y.A., Reed,S.G. and Cheever,M.A.
TITLE Compositions and methods for therapy and diagnosis of prostate
cancer
JOURNAL Patent: WO 0125272-A 183 12-APR-2001;
CORIXA CORPORATION (US)
FEATURES
source
1. .384
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 102 a 102 c 90 g 89 t
ORIGIN

Query Match 96.2% Score 367.4; DB 6; Length 384;
Best Local Similarity 99.0%; Pred. No. 6.1e-96;
Matches 379; Conservative 1; Mismatches 2; Indels 1; Gaps 1;

Oy 1 aggcggagcagaagctaaagccaaagcccaag-agaagtggcagtgccagcactggcgcc 59
Db 1 AGCGGGAGCAGAGAGCTAAAGCCAAAGCCCAAGAGTGCGCAGTGCCAGCCTGGTGCC 50
Oy 60 agtaccagtaccataaaccagtgccagtgccagtgccagcaccagtggtggttcagtgct 119
Db 61 AGTACCAGTACCATAACAGTGGCAGTGCCAGTGCCAGTGCCAGCACCAGTGGTGGCTTCAGTGCT 120
Oy 120 ggtgcagcctgacccgcaactctcacatttgggctcttcgtggtccttggtgagcgtggt 179
Db 121 GGTGCCAGCCTGACCGCCACTCTCACATTTGGGCTCTTCGCTGGCCTTGGTGAGCTGGT 180
Oy 180 gccagcaccagtgagcagctctggtcctggttctcctcacaagtgagatttagatat 239
Db 181 GCCAGCACCAGTGGCAGCTCTGGTGGCTTGTGTTTCTTCTTCAAGTGTGAGATTTTATGATAT 240
Oy 240 tattaactctgcagctcttctccttaagccagggtgcatcctcagaacacctactcaaca 299
Db 241 TGTTAATCTGCCAGCTCTTCTTCTTCAAGCCAGGGTGCATCTCAGAACCTTACTTCAACA 300
Oy 300 cagcactctaggcagccactatcaatcaattgaagttagacactctgcattaaatctattt 359
Db 301 CAGCAGCTCTAGCAGCCACTATCAATCAATTGAAGTTGACACTCTGCAITTAATCTATT 360
Oy 360 gccattataaaaaaa 382
Db 361 GCAATTTCAAAAAA 383

RESULT 7
AX140693
LOCUS AX140693 384 bp DNA linear PAT 31-MAY-2001
DEFINITION Sequence 183 from Patent WO0134802.
ACCESSION AX140693
VERSION AX140693.1 GI:14280804
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 384)
AUTHORS Xu,J., Dillon,D.C., Mitcham,J.L., Harlocker,S.L., Jiang,Y.,
Reed,S.G., Kalos,M.D., Reltter,M.W., Stolk,J.A., Day,C.H.,
```

Skeiky,Y.A. and Wang,A.  
Compositions and methods for the therapy and diagnosis of prostate cancer

Patent: WO 0134802-A 183 17-MAY-2001;

CORIXA CORPORATION (US)

Location/Qualifiers

FEATURES

source

1. 384  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"

BASE COUNT 102 a 102 c 90 g 89 t 1 others  
ORIGIN

Query Match 95.2%; Score 367.4; DB 6; Length 384;  
Best Local Similarity 99.0%; Pred. No. 6.le-96;  
Matches 379; Conservative 1; Mismatches 2; Indels 1; Gaps 1;

Qy 1 agcgggagcagaagctaaagccaaagcccaag-agaagtggcagtgccagcactggtgcc 59  
|||||  
Db 1 AGCGGGAGCAGAAAGCTAAAGCCAAAGCCCAAGAGAGTGGCAGTGCCAGCAGTGGTGC 60

Qy 60 agtaccagctaccaataaacagtgccagtgccagtgccagcagcagtggtgcagtgct 119  
|||||  
Db 61 AGTACCAGTACCAATAACAGTGCCAGTGCCAGTGCCAGTGCCAGTGCCAGTGCT 120

Qy 120 ggtgccagctgacccgacctctcacatttgggctcttcgctggccttggctggagctgct 179  
|||||  
Db 121 GGTGCCAGCTGACCGCCACTCTCACATTTGGGCTCTTCGCTGGCTTGGTGGAGCTGT 180

Qy 180 gccagcacttagcagcagcactatcaattgaagtgacacttgacacttgacataatctatt 239  
|||||  
Db 181 GCCAGCAGTGGCAGCTCTGGTGGCTTGGTGGCTTGGTGGCTTGGTGGCTTGGTGGCT 240

Qy 240 ttttaactctgccagctcttcttcaagcaggtgcatctcagaaacctactcaaca 299  
|||||  
Db 241 TGTAACTCTGCCAGTCTTCTTCAAGCCAGGTGTCATCTCAGAAACCTACTCAACA 300

Qy 300 cagcacttagcagcagcactatcaattgaagtgacacttgacacttgacataatctatt 359  
|||||  
Db 301 CAGCACTCTAGCGAGCCACTATCAATCAATTTGAAGTTGACACTCTGCATTTATATCTATT 360

Qy 360 gccattataaaaaa 382  
|||||  
Db 361 GCCATTTCAAAAA 383

RESULT 8  
AX200553  
LOCUS AX200553 384 bp DNA linear PAI 23-AUG-2001  
DEFINITION Sequence 183 from Patent WO0151633.  
ACCESSION AX200553  
VERSION AX200553.1 GI:15390369  
KEYWORDS human.  
SOURCE Homo sapiens  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Euthera; Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE 1 (bases 1 to 384)  
AUTHORS Xu,J., Dillon,D.C., Mitcham,J.L., Harlocker,S.L., Jiang,Y.,  
Reed,S.G., Kalos,M.D., Fanger,G.R., Retter,M.W., Day,C.H., Retter,M.W.,  
Stolk,J.A., Skeiky,Y.A., Wang,A. and Meagher,M.J.  
Compositions and methods for the therapy and diagnosis of prostate cancer  
Patent: WO 0151633-A 183 19-JUL-2001;  
CORIXA CORPORATION (US)  
Location/Qualifiers  
1. 384  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"

BASE COUNT 102 a 102 c 90 g 89 t 1 others  
ORIGIN

Query Match 96.2%; Score 367.4; DB 6; Length 384;  
Best Local Similarity 99.0%; Pred. No. 6.le-96;  
Matches 379; Conservative 1; Mismatches 2; Indels 1; Gaps 1;

Qy 1 agcgggagcagaagctaaagccaaagcccaag-agaagtggcagtgccagcactggtgcc 59  
|||||  
Db 1 AGCGGGAGCAGAAAGCTAAAGCCAAAGCCCAAGAGAGTGGCAGTGCCAGCAGTGGTGC 60

Qy 60 agtaccagctaccaataaacagtgccagtgccagtgccagcagcagtggtgcagtgct 119  
|||||  
Db 61 AGTACCAGTACCAATAACAGTGCCAGTGCCAGTGCCAGTGCCAGTGCTTTCAGTGCT 120

Qy 120 ggtgccagctgacccgacctctcacatttgggctcttcgctggccttggctggagctgct 179  
|||||  
Db 121 GGTGCCAGCTGACCGCCACTCTCACATTTGGGCTCTTCGCTGGCTTGGTGGAGCTGT 180

Qy 180 gccagcacttagcagcagcactatcaattgaagtgacacttgacacttgacataatctatt 239  
|||||  
Db 181 GCCAGCAGTGGCAGCTCTGGTGGCTTGGTGGCTTGGTGGCTTGGTGGCTTGGTGGCT 240

Qy 240 ttttaactctgccagctcttcttcaagcaggtgcatctcagaaacctactcaaca 299  
|||||  
Db 241 TGTAACTCTGCCAGTCTTCTTCAAGCCAGGTGTCATCTCAGAAACCTACTCAACA 300

Qy 300 cagcacttagcagcagcactatcaattgaagtgacacttgacacttgacataatctatt 359  
|||||  
Db 301 CAGCACTCTAGCGAGCCACTATCAATCAATTTGAAGTTGACACTCTGCATTTATATCTATT 360

Qy 360 gccattataaaaaa 382  
|||||  
Db 361 GCCATTTCAAAAA 383

RESULT 9  
AX267209  
LOCUS AX267209 384 bp DNA linear PAT 26-OCT-2001  
DEFINITION Sequence 183 from Patent WO0173032.  
ACCESSION AX267209  
VERSION AX267209.1 GI:16515987  
KEYWORDS human.  
SOURCE Homo sapiens  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Euthera; Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE 1 (sites)  
AUTHORS Xu,J., Dillon,D.C., Mitcham,J.L., Harlocker,S.L., Jiang,Y.,  
Kalos,M.D., Fanger,G.R., Retter,M.W., Stolk,J.A., Day,C.H.,  
Vedvick,I.S., Carter,D., Li,S.X., Wang,A., Skeiky,Y.A., Hepler,W.I.  
and Henderson,R.A.  
Compositions and methods for the therapy and diagnosis of prostate cancer  
Patent: WO 0173032-A 183 04-OCT-2001;  
CORIXA CORPORATION (US)  
Location/Qualifiers  
1. 384  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"

BASE COUNT 102 a 102 c 90 g 89 t 1 others  
ORIGIN

Query Match 96.2%; Score 367.4; DB 6; Length 384;  
Best Local Similarity 99.0%; Pred. No. 6.le-96;  
Matches 379; Conservative 1; Mismatches 2; Indels 1; Gaps 1;

Qy 1 agcgggagcagaagctaaagccaaagcccaag-agaagtggcagtgccagcactggtgcc 59  
|||||  
Db 1 AGCGGGAGCAGAAAGCTAAAGCCAAAGCCCAAGAGAGTGGCAGTGCCAGCAGTGGTGC 60

Qy 60 agtaccagctaccaataaacagtgccagtgccagtgccagcagcagtggtgcagtgct 119  
|||||  
Db 61 AGTACCAGTACCAATAACAGTGCCAGTGCCAGTGCCAGTGCCAGTGCTTTCAGTGCT 120

```
Qy 120 ggtgcccagcctgaccgcccactctcaccatttggtgctctcgctggccttgctgagctggt 179
|||||
Db 121 ggtgcccagcctgaccgcccactctcaccatttggtgctctcgctggccttgctgagctggt 180
|||||
Qy 180 gccagcaccagtgagcagctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 239
|||||
Db 181 gccagcaccagtgagcagctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 240
|||||
Qy 240 tggtaatcctgcccagctcttcttcttcaagccaggggtgctcctcctcagaaacctactcaaca 299
|||||
Db 241 tggtaatcctgcccagctcttcttcttcaagccaggggtgctcctcctcagaaacctactcaaca 300
|||||
Qy 300 cagcactctaggcagcactatcaatcaattgaagttgacactctgacattcaatctattt 359
|||||
Db 301 cagcactctaggcagcactatcaatcaattgaagttgacactctgacattcaatctattt 360
|||||
Qy 360 gccattataaaaaa 382
|||||
Db 361 gccattataaaaaa 382
|||||

RESULT 10
LOCUS AX106301 383 bp DNA linear PAT 30-APR-2001
DEFINITION Sequence 82 from Patent WO0125272.
ACCESSION AX106301
VERSION AX106301.1 GI:13921987
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Xu,J., Skeiky,Y.A., Reed,S.G. and Cheever,M.A.
TITLE Compositions and methods for therapy and diagnosis of prostate
cancer
JOURNAL Patent: WO 0125272-A 82 12-APR-2001;
CORIXA CORPORATION (US)
FEATURES
Source
misc_feature 1..383
/db_xref="taxon:9606"
BASE COUNT 102 a 102 c 89 g 89 t 1 others
ORIGIN
Query Match 92.9%; Score 354.8; DB 5; Length 383;
Best Local Similarity 98.7%; Pred. No. 2.8e-92;
Matches 378; Conservative 0; Mismatches 3; Indels 2; Gaps 2:

Qy 1 agcgggagcagaagcctaaagccaaagcccaag-aggatggcagtgccagcactggtgcc 59
|||||
Db 1 agcgggagcagaagcctaaagccaaagcccaag-aggatggcagtgccagcactggtgcc 50
|||||
Qy 60 agtaccagttaccaataacagtcagctgcccagtcagtcagtcagtcagtcagtcagtcagtc 119
|||||
Db 61 agtaccagttaccaataacagtcagctgcccagtcagtcagtcagtcagtcagtcagtcagtc 119
|||||
Qy 120 ggtgcccagcctgaccgcccactctcaccatttggtgctctcgctggccttggtgagctggt 179
|||||
Db 120 ggtgcccagcctgaccgcccactctcaccatttggtgctctcgctggccttggtgagctggt 179
|||||
Qy 180 gccagcaccagtgagcagctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 239
|||||
Db 180 gccagcaccagtgagcagctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 239
|||||
Qy 240 tggtaatcctgcccagctcttcttcttcaagccaggggtgctcctcctcagaaacctactcaaca 299
|||||
Db 240 tggtaatcctgcccagctcttcttcttcaagccaggggtgctcctcctcagaaacctactcaaca 299
|||||
Qy 300 cagcactctaggcagcactatcaatcaattgaagttgacactctgacattcaatctattt 359
|||||
Db 300 cagcactctaggcagcactatcaatcaattgaagttgacactctgacattcaatctattt 359
|||||
Qy 360 gccattataaaaaa 382
|||||
Db 360 gccattataaaaaa 382
|||||

RESULT 12
LOCUS AX200452 383 bp DNA linear PAT 29-AUG-2001
DEFINITION Sequence 82 from Patent WO0151633.
ACCESSION AX200452
```

```
Db 300 cagcactctaggcagcactatcaatcaattgaagttgacactctgacattcaatctattt 359
|||||
Qy 360 gccattataaaaaa 382
|||||
Db 360 gccattataaaaaa 382
|||||

RESULT 11
LOCUS AX140592 383 bp DNA linear PAT 31-MAY-2001
DEFINITION Sequence 82 from Patent WO0134802.
ACCESSION AX140592
VERSION AX140592.1 GI:14280710
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Xu,J., Dillon,D.C., Mitcham,J.L., Harlocker,S.L., Jiang,Y.,
Reed,S.G., Kalos,M.D., Retter,M.W., Stolk,J.A., Day,C.H.,
Skeiky,Y.A. and Wang,A.
TITLE Compositions and methods for the therapy and diagnosis of prostate
cancer
JOURNAL Patent: WO 0134802-A 82 17-MAY-2001;
CORIXA CORPORATION (US)
FEATURES
Source
misc_feature 1..383
/db_xref="taxon:9606"
BASE COUNT 102 a 102 c 89 g 89 t 1 others
ORIGIN
Query Match 92.9%; Score 354.8; DB 6; Length 383;
Best Local Similarity 98.7%; Pred. No. 2.8e-92;
Matches 378; Conservative 0; Mismatches 3; Indels 2; Gaps 2:

Qy 1 agcgggagcagaagcctaaagccaaagcccaag-aggatggcagtgccagcactggtgcc 59
|||||
Db 1 agcgggagcagaagcctaaagccaaagcccaag-aggatggcagtgccagcactggtgcc 50
|||||
Qy 60 agtaccagttaccaataacagtcagctgcccagtcagtcagtcagtcagtcagtcagtcagtc 119
|||||
Db 61 agtaccagttaccaataacagtcagctgcccagtcagtcagtcagtcagtcagtcagtcagtc 119
|||||
Qy 120 ggtgcccagcctgaccgcccactctcaccatttggtgctctcgctggccttggtgagctggt 179
|||||
Db 120 ggtgcccagcctgaccgcccactctcaccatttggtgctctcgctggccttggtgagctggt 179
|||||
Qy 180 gccagcaccagtgagcagctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 239
|||||
Db 180 gccagcaccagtgagcagctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 239
|||||
Qy 240 tggtaatcctgcccagctcttcttcttcaagccaggggtgctcctcctcagaaacctactcaaca 299
|||||
Db 240 tggtaatcctgcccagctcttcttcttcaagccaggggtgctcctcctcagaaacctactcaaca 299
|||||
Qy 300 cagcactctaggcagcactatcaatcaattgaagttgacactctgacattcaatctattt 359
|||||
Db 300 cagcactctaggcagcactatcaatcaattgaagttgacactctgacattcaatctattt 359
|||||
Qy 360 gccattataaaaaa 382
|||||
Db 360 gccattataaaaaa 382
|||||

RESULT 12
LOCUS AX200452 383 bp DNA linear PAT 29-AUG-2001
DEFINITION Sequence 82 from Patent WO0151633.
ACCESSION AX200452
```

```

VERSION      AX200452.1  GI:15390261
KEYWORDS
SOURCE       human.
ORGANISM     Homo sapiens
REFERENCE    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS      Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
              1 (bases 1 to 383)
              Xu, J., Dillon, D.C., Mitcham, J.L., Harlocker, S.L., Jiang, Y.,
              Peed, S.G., Kalos, M.D., Fanger, G.R., Day, C.H., Retter, M.W.,
              Stolk, J.A., Skeiky, Y.A., Wang, A., and Meagher, M.J.
              Stolk, J.A., Skeiky, Y.A., Wang, A., and Meagher, M.J.
              Compositions and methods for the therapy and diagnosis of prostate
              cancer
              Patent: WO 0151633-A 82 19-JUL-2001;
              CORIXA CORPORATION (US)
              Location/Qualifiers
              1..383
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              102 a 102 c 89 g 89 t 1 others
              BASE COUNT 102 a 102 c 89 g 89 t 1 others
              ORIGIN
              Query Match 92.9%; Score 354.8; DB 5; Length 383;
              Best Local Similarity 98.7%; Pred. No. 2.8e-92;
              Matches 378; Conservative 0; Mismatches 3; Indels 2; Gaps 2;
              QY 1 agcgaggagcagaagctaaagccaaagccaaag-agaagtgccagtgccagcactgagtgcc 59
              Db 1 AGCGGGAGCAGAGCTAAAGCCAAAGCCAAAGAGAGTGGCAGTGGCAGCACITGGTGCC 60
              QY 60 agtaccagtaaccaataaacagtgccagtgccagtgccagcaccagtggtggttcagtgct 119
              Db 61 AGTACCAGTACCAATAACA-TGCCAGTGCCAGTGCCAGCAGCAGTGCGTTCAGTGTCT 119
              QY 120 ggtgccagctgaccgcccactctccacatttgggctctcgtgccccttggaggagctggt 179
              Db 120 GGTGCCAGCTGACCCGCCACTCTCACATTTGGGCTCTTGGCTGGCTTGGTGAGAGCTGT 179
              QY 180 gccagcaccagtgccagctcagtgccctggtggttctctcaagtgagatttagatat 239
              Db 180 GCCAGCACCAGTGGCAGCTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGT 239
              QY 240 tggtaacctgcccagctcttctcttctcaagcgggtgagctctcagaaacctactcaaca 299
              Db 240 TGTTAATCTGCGCAGCTCTTCTCTTCAAGCCAGGGTGCACTCTCAGAAACCTACTCAACA 299
              QY 300 cagcactctgagcagcactatcaatcaatgaagtgcacactctgcattataatctatt 359
              Db 300 CAGCAGCTCTGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 359
              QY 360 gccattaaaaaataaaaaa 382
              Db 360 GCCATTTCAAAAAATAAAAAA 382
              RESULT 13
              AX267108 383 bp DNA linear PAT 25-OCT-2001
              LOCUS Sequence 82 from Patent WO0173032.
              DEFINITION AX267108
              ACCESSION AX267108
              VERSION AX267108.1 GI:15515893
              KEYWORDS
              SOURCE human.
              ORGANISM Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
              1 (sites)
              Xu, J., Dillon, D.C., Mitcham, J.L., Harlocker, S.L., Jiang, Y.,
              Kalos, M.D., Fanger, G.R., Retter, M.W., Stolk, J.A., Day, C.H.,
              Vedvick, I.S., Carter, D., Li, S.X., Wang, A., Skeiky, Y.A., Hepler, W.T.
              and Henderson, R.A.
              Compositions and methods for the therapy and diagnosis of prostate
              cancer
              TITLE
    
```

```

JOURNAL      Patent: WO 0173032-A 82 04-OCT-2001;
              CORIXA CORPORATION (US)
              Location/Qualifiers
              1..383
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              102 a 102 c 89 g 89 t 1 others
              BASE COUNT 102 a 102 c 89 g 89 t 1 others
              ORIGIN
              Query Match 92.9%; Score 354.8; DB 6; Length 383;
              Best Local Similarity 98.7%; Pred. No. 2.8e-92;
              Matches 378; Conservative 0; Mismatches 3; Indels 2; Gaps 2;
              QY 1 agcgaggagcagaagctaaagccaaagccaaag-agaagtgccagtgccagcactgagtgcc 59
              Db 1 AGCGGGAGCAGAGCTAAAGCCAAAGCCAAAGAGAGTGGCAGTGGCAGCACITGGTGCC 60
              QY 60 agtaccagtaaccaataaacagtgccagtgccagtgccagcaccagtggtggttcagtgct 119
              Db 61 AGTACCAGTACCAATAACA-TGCCAGTGCCAGTGCCAGCAGCAGTGCGTTCAGTGTCT 119
              QY 120 ggtgccagctgaccgcccactctccacatttgggctctcgtgccccttggaggagctggt 179
              Db 120 GGTGCCAGCTGACCCGCCACTCTCACATTTGGGCTCTTGGCTGGCTTGGTGAGAGCTGT 179
              QY 180 gccagcaccagtgccagctcagtgccctggtggttctctcaagtgagatttagatat 239
              Db 180 GCCAGCACCAGTGGCAGCTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGT 239
              QY 240 tggtaacctgcccagctcttctcttctcaagcgggtgagctctcagaaacctactcaaca 299
              Db 240 TGTTAATCTGCGCAGCTCTTCTCTTCAAGCCAGGGTGCACTCTCAGAAACCTACTCAACA 299
              QY 300 cagcactctgagcagcactatcaatcaatgaagtgcacactctgcattataatctatt 359
              Db 300 CAGCAGCTCTGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 359
              QY 360 gccattaaaaaataaaaaa 382
              Db 360 GCCATTTCAAAAAATAAAAAA 382
              RESULT 14
              AF128527 2001 bp mRNA linear PRI 01-JUN-1999
              LOCUS Homo sapiens clone snrl breast cancer associated gene 1 protein
              DEFINITION (BCG1) mRNA, complete cds.
              ACCESSION AF128527
              VERSION AF128527.1 GI:4928043
              KEYWORDS
              SOURCE human.
              ORGANISM Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
              1 (bases 1 to 2001)
              Kurt, R.A. and Urban, W.J.
              Isolation of breast cancer associated gene 1 from normal mammary
              epithelium
              unpublished
              2 (bases 1 to 2001)
              Kurt, R.A. and Urban, W.J.
              Direct Submission
              Submitted (15-FEB-1999) Cellular Immunology, Earle A. Childs
              Research Institute, Franz Cancer Research Center, 4805 NE Glisan,
              Portland, OR 97213, USA
              Location/Qualifiers
              1..2001
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              /clone="snrl"
              /tissue_type="normal mammary epithelium"
              1..2001
              gene
    
```

	BASE COUNT	521 a	516 c	569 g	395 t	ORIGIN	
CDS		/gene="BCG1" 70..1890 /gene="BCG1" /codon_start=1 /product="breast cancer associated gene 1 protein" /protein_id="AAD33392.1" /db_xref="GI:4928044" /translation="MSDTESGAGLTRFOAESEKDSMMOITLLTIVONVEVPEPK ASKALEVEDSVKSAQSKATKYEVSIPKPEAREAPATQASSITQLTDTQVLAENKSL AADTKKQADPQVAVTAPATETKTKVSHVADIRVYIKQETEAAPSOAPDEPPESAA QSQENQDTRPKVKAKKARKKHLDEEDGSDSQASGTTGGRRVSKATLWASWARRAS RQPIAFWARRASRTPLAWARRALLSLRSPKARRKARRAAKQLQSRPEAPPDPRV ALLOGENDVLYLLAKQDTIKPKRMDLKDKIKEYTDVYPIEIIENAGRSLEKVG ELKIDKNDHLYLLASTLEPTDAGILGTIKDPSKLGLMLVLSIIIFMGNSSEAVIN EVLRKLGRLPGHIHSFGDGVKLLITDFVKQYLDYVARFVNSNPPEYFEFFWGLRGLG I SRMKVLRFCAGKQKDKPKNWALRYDEAMEADLKAAAEAAAKARAEIIPARMGILG SENAGCPNWDIEDIGWAKARKIOAGAEAKAKAQSASGASTGASISTHNSASASASTS GFSGASGLATLITGLFAGLUGAGAGATSGSGAGCGFSYK"					
		521 a	516 c	569 g	395 t		
	Query Match	98.5%	Score	338	DB 9:	Length 2001:	
	Best Local Similarity	100.0%	Pred. No.	2.7e-87:			
	Matches 338:	Conservative	0:	Mismatches	0:	Indels	
					0:	Gaps	
Oy	1	agcgggagcagaagactaaagccaaagcccaagagagtggcagtcgacacacttgatgcca	50				
Db	1664	AGGCGGGAGCAGAGCTAAGCCAAAGCCCAAGAGAGTGGCAGTGCAGCACTGGTGCCA	1723				
Oy	61	gtaccagttaccaataaacaagtcaccagtcaccagtcaccagtcacagtcgtgcttcagtcgtg	120				
Db	1724	GTACCAGTACC AAT AACAGT GCCAGT GCCAGT GCCAGCACCAGTGGTGGCTTCAGTGTG	1783				
Oy	121	gtgccagtcctgaccgccactctcacatttggtctcttcgtgcaccttggtgaagctcggctg	180				
Db	1784	GTGCCAGCTTGACCCGACICTCACATITGGGCTCTTCGCTGGCCITTTGGTGGAGCTGGTG	1843				
Oy	181	ccagccaccagtcggcagctcttgctgccttgattctctcacaagtcagattttagatatt	240				
Db	1844	CCAGCACCAGTGGCAGCTCTGTGGCTCTGGTTTCTCTCTACAGTGGAGATTTTAGATATT	1903				
Oy	241	gttaatccgcagctcttctcttcaagccagggtgcatactctcagaacactactcaaac	300				
Db	1904	GTTAATCTTCCGACGTCTTCTCTCAAGCCAGGTGTCATCTCTCAGAAACCTTACTCAACAC	1963				
Oy	301	agcacttagcgagccactatcaatcaattgaattga	338				
Db	1964	AGCACTTAGCGAGCCACTATCAATCAATTGAAGTTGA	2001				

RESULT	15
AXI05292	JLOCUS
AXI06292	Sequence 73 from Patent WO0125272.
AXI06292	ACCESSION
AXI06292.1	GI:13221978
AXI06292.1	VERSION
KEYWORDS	human.
SOURCE	Homo sapiens
ORGANISM	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE	1 (bases 1 to 499)
AUTHORS	Xu,J.,Skeiky,Y.A., Peed,S.G. and Cheever,M.A.
TITLE	Compositions and methods for therapy and diagnosis of prostate cancer
JOURNAL	Patent: WO 0125272-A 73 12-APR-2001:
FEATURES	CORIXA CORPORATION (US)
source	Location/Qualifiers
	1..499
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	1..499
misc	feature

1000000



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic --nucleic search, using sw model

Run on: April 30, 2002, 10:36:25 ; Search time 308.52 Seconds  
(without alignments)  
1004.220 Million cell updates/sec

Title: US-09-248-178-61  
Perfect score: 1368  
Sequence: 1 ccagtgagcgcgtaatac.....ccgcttgagcgaatacatnn 1368

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 351203 seqs, 113238999 residues

Word size : 10

Total number of hits satisfying chosen parameters: 107588

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued\_Patents\_NA.\*  
1: /cgn2\_6ptodata/1/ina/5A\_COMB.seq.\*  
2: /cgn2\_6ptodata/1/ina/5B\_COMB.seq.\*  
3: /cgn2\_6ptodata/1/ina/6A\_COMB.seq.\*  
4: /cgn2\_6ptodata/1/ina/6B\_COMB.seq.\*  
5: /cgn2\_6ptodata/1/ina/PTUS\_COMB.seq.\*  
6: /cgn2\_6ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	68	5.0	703	4	US-08-998-416-136 Sequence 136, Appl
2	68	5.0	716	4	US-08-998-416-55 Sequence 55, Appl
3	68	5.0	793	4	US-09-020-956-33 Sequence 33, Appl
4	68	5.0	793	4	US-09-030-607-33 Sequence 33, Appl
5	68	5.0	801	2	US-08-975-316-50 Sequence 50, Appl
6	68	5.0	814	4	US-09-020-956-1 Sequence 1, Appl
7	68	5.0	814	4	US-09-030-607-1 Sequence 1, Appl
8	68	5.0	816	4	US-09-020-956-2 Sequence 2, Appl
9	68	5.0	816	4	US-09-030-607-2 Sequence 2, Appl
10	68	5.0	949	2	US-08-713-000-4 Sequence 4, Appl
11	68	5.0	949	2	US-08-975-316-4 Sequence 4, Appl
12	68	5.0	949	2	US-09-211-710-4 Sequence 4, Appl
13	68	5.0	2961	4	US-08-446-935-6 Sequence 6, Appl
14	68	5.0	3792	2	US-08-992-334-1 Sequence 1, Appl
15	68	5.0	3792	2	US-08-302-752-1 Sequence 1, Appl
16	68	5.0	4145	3	US-08-651-472-52 Sequence 52, Appl
17	68	5.0	4145	3	US-08-358-928-52 Sequence 52, Appl
18	68	5.0	4277	3	US-08-651-472-63 Sequence 63, Appl
19	68	5.0	4277	3	US-08-358-928-63 Sequence 63, Appl
20	68	5.0	5234	2	US-08-992-334-2 Sequence 2, Appl
21	68	5.0	5234	3	US-08-302-752-2 Sequence 2, Appl
22	68	5.0	5356	4	US-08-446-935-1 Sequence 1, Appl
23	68	5.0	5532	3	US-08-651-472-72 Sequence 72, Appl
24	68	5.0	5532	3	US-08-358-928-72 Sequence 72, Appl
25	68	5.0	6722	2	US-08-992-334-3 Sequence 3, Appl
26	68	5.0	6722	3	US-08-302-752-3 Sequence 3, Appl
27	68	5.0	6811	3	US-08-651-472-67 Sequence 67, Appl

c 28	68	5.0	6811	4	US-08-358-928-67	Sequence 67, Appl
c 29	68	5.0	6926	3	US-08-651-472-69	Sequence 69, Appl
c 30	68	5.0	6926	4	US-08-358-928-69	Sequence 69, Appl
c 31	62	4.5	558	4	US-08-930-285-17	Sequence 17, Appl
c 32	52	4.5	698	4	US-08-998-416-54	Sequence 54, Appl
c 33	62	4.5	713	4	US-08-998-416-135	Sequence 135, Appl
c 34	62	4.5	5534	1	US-08-452-267-3	Sequence 3, Appl
c 35	62	4.5	5534	3	US-09-123-644-3	Sequence 3, Appl
c 36	62	4.5	7659	4	US-09-128-314-4	Sequence 4, Appl
c 37	62	4.5	8157	4	US-09-128-314-3	Sequence 3, Appl
c 38	61	4.5	2363	1	US-08-098-947-2	Sequence 2, Appl
c 39	61	4.5	2363	1	US-07-919-140B-2	Sequence 2, Appl
c 40	61	4.5	2363	4	US-08-916-232-2	Sequence 2, Appl
c 41	61	4.5	2363	5	PCT-US93-06939-2	Sequence 2, Appl
c 42	55	4.0	5618	3	US-08-799-569-1	Sequence 1, Appl
c 43	51	3.7	329	1	US-08-703-809-9	Sequence 9, Appl
c 44	51	3.7	329	1	US-08-703-808-9	Sequence 9, Appl
c 45	51	3.7	329	2	US-08-703-807-9	Sequence 9, Appl

ALIGNMENTS

RESULT 1  
US-08-998-416-136  
: Sequence 136, Application US/08998416  
: Patent No. 6239264  
: GENERAL INFORMATION:  
: APPLICANT: Philippsen, Peter  
: APPLICANT: Pohlmann, Rainer  
: APPLICANT: Steiner, Sabine  
: APPLICANT: Mohr, Christine  
: APPLICANT: Wendland, Jurgen  
: APPLICANT: Knechtie, Philipp  
: APPLICANT: Rebeschung, Corinne  
: TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSYPII  
: TITLE OF INVENTION: AND USES THEREOF  
: NUMBER OF SEQUENCES: 1152  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: No. 6239264artis Corporation  
: STREET: 3054 Cornwallis Road  
: CITY: Research Triangle Park  
: STATE: No. 6239264th Carolina  
: COUNTRY: USA  
: ZIP: 27709  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: Floppy disk  
: COMPUTER: IBM PC compatible  
: OPERATING SYSTEM: PC-DOS/MS-DOS  
: SOFTWARE: PatentIn Release #1.0, Version #1.30  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/08/998,416  
: FILING DATE: 24-DEC-1997  
: CLASSIFICATION: 435  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: CH 0016/97  
: FILING DATE: 31-DEC-1996  
: ATTORNEY/AGENT INFORMATION:  
: NAME: Meigs, J. Timothy  
: REGISTRATION NUMBER: 38,241  
: REFERENCE/DOCKET NUMBER: PF/5-30306/A/GC1976  
: TELECOMMUNICATION INFORMATION:  
: TELEPHONE: 919-541-8587  
: TELEFAX: 919-541-8689  
: INFORMATION FOR SEQ ID NO: 136:  
: SEQUENCE CHARACTERISTICS:  
: LENGTH: 703 base pairs  
: TYPE: nucleic acid  
: STRANDEDNESS: single  
: TOPOLOGY: linear  
: MOLECULE TYPE: DNA (genomic)  
: ORIGINAL SOURCE:  
: ORGANISM: PAG1051UP



```

CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/030.607
FILING DATE: 25-FEB-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.427C3
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 33:
SEQUENCE CHARACTERISTICS:
LENGTH: 793 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cdna
US-09-030-607-33

```

```

Query Match          5.0%; Score 68; DB 4; Length 793;
Best Local Similarity 100.0%; Pred. No. 8.6e-18;
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1299 gccaccgcgggtgagctccagctttgtcccttttagtgagggttaattgcgcgttgc 1358
|||||
Db 425 gccaccgcgggtgagctccagctttgtcccttttagtgagggttaattgcgcgttgc 484
|||||

Oy 1359 gtaatcat 1365
|||||
Db 485 GTAATCAT 492

RESULT 5
US-08-975-316-50/c
Sequence 50, Application US/08975316
Patent No. 5952486
GENERAL INFORMATION:
APPLICANT: BLOKSBERG, Leonard N., HAVUKKALA, Ilkka
APPLICANT: and GRIERSON, Alastair W.
TITLE OF INVENTION: MATERIALS AND METHODS FOR
TITLE OF INVENTION: THE MODIFICATION OF PLANT LIGNIN CONTENT
NUMBER OF SEQUENCES: 88
CORRESPONDENCE ADDRESS:
ADDRESSEE: Law Offices of Ann W. Speckman
STREET: 2601 Elliott Avenue, Suite 4185
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98121
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/975,315
FILING DATE:
CLASSIFICATION: 800
PRIOR APPLICATION DATA:

```

```

APPLICATION NUMBER: 08/713,000
FILING DATE: September 11, 1996
ATTORNEY/AGENT INFORMATION:
NAME: SLEATH, Janet
REGISTRATION NUMBER: 37,007
REFERENCE/DOCKET NUMBER: 11000/1003C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-269-0565
TELEFAX: 206-269-0563
TELEX:
INFORMATION FOR SEQ ID NO: 50:
SEQUENCE CHARACTERISTICS:
LENGTH: 801 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-975-316-50

```

```

Query Match          5.0%; Score 68; DB 2; Length 801;
Best Local Similarity 100.0%; Pred. No. 8.6e-18;
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Oy 1299 gccaccgcgggtgagctccagctttgtcccttttagtgagggttaattgcgcgttgc 1358
|||||
Db 246 gccaccgcgggtgagctccagctttgtcccttttagtgagggttaattgcgcgttgc 187
|||||

```

```

Oy 1359 gtaatcat 1366
|||||
Db 186 GTAATCAT 179

```

```

RESULT 5
US-09-020-956-1
Sequence 1, Application US/09020956
Patent No. 6261562
GENERAL INFORMATION:
APPLICANT: Xu, Jiangchun
APPLICANT: Dillin, David C.
TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE CANCER AND METHODS
NUMBER OF SEQUENCES: 178
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/020,956
FILING DATE: 09-FEB-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.427C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 814 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cdna
US-09-020-956-1

```

```
Query Match          5.0%: Score 68; DB 4: Length 814;
Best Local Similarity 100.0%: Pred. No. 8.6e-18;
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1299 gccaccgcggtgagctccagcttttgccttttagtgagggttaattgcgcgttggc 1358
|||||
Db 372 GCCACCGCGGTGAGCTCCAGCTTTTGTTCCTTTAGTGAGGGTTAATTGCGCGCTTGGC 431

OY 1359 gtaatcat 1366
|||||
Db 432 GTAATCAT 439

RESULT 7
US-09-030-607-1
: Sequence 1, Application US/09030607
: Patent No. 6262245
: GENERAL INFORMATION:
: APPLICANT: Xu, Jiangchun
: TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE CANCER AND METHODS FO
: NUMBER OF SEQUENCES: 224
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: SEED and BERRY LLP
: STREET: 6300 Columbia Center, 701 Fifth Avenue
: CITY: Seattle
: STATE: WA
: COUNTRY: USA
: ZIP: 98104
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: FILING DATE: 25-FEB-1998
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: Maki, David J.
: REGISTRATION NUMBER: 31,392
: REFERENCE/DOCKET NUMBER: 210121.427C3
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (206) 622-4900
: TELEFAX: (206) 682-6031
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 814 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cdna
US-09-030-607-1

Query Match          5.0%: Score 68; DB 4: Length 814;
Best Local Similarity 100.0%: Pred. No. 8.6e-18;
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1299 gccaccgcggtgagctccagcttttgccttttagtgagggttaattgcgcgttggc 1358
|||||
Db 372 GCCACCGCGGTGAGCTCCAGCTTTTGTTCCTTTAGTGAGGGTTAATTGCGCGCTTGGC 431

OY 1359 gtaatcat 1366
|||||
Db 432 GTAATCAT 439

RESULT 8
US-09-020-956-2
: Sequence 2, Application US/09020956
: Patent No. 6261562
```

```
GENERAL INFORMATION:
: APPLICANT: Xu, Jiangchun
: TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE CANCER AND METHODS
: NUMBER OF SEQUENCES: 178
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: SEED and BERRY LLP
: STREET: 6300 Columbia Center, 701 Fifth Avenue
: CITY: Seattle
: STATE: WA
: COUNTRY: USA
: ZIP: 98104
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: FILING DATE: 09-FEB-1998
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: Maki, David J.
: REGISTRATION NUMBER: 31,392
: REFERENCE/DOCKET NUMBER: 210121.427C2
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (206) 622-4900
: TELEFAX: (206) 682-6031
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 815 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cdna
US-09-020-956-2

Query Match          5.0%: Score 68; DB 4: Length 816;
Best Local Similarity 100.0%: Pred. No. 8.6e-18;
Matches 58; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1299 gccaccgcggtgagctccagcttttgccttttagtgagggttaattgcgcgttggc 1358
|||||
Db 474 GCCACCGCGGTGAGCTCCAGCTTTTGTTCCTTTAGTGAGGGTTAATTGCGCGCTTGGC 483

OY 1359 gtaatcat 1366
|||||
Db 484 GTAATCAT 491

RESULT 9
US-09-030-507-2
: Sequence 2, Application US/09030607
: Patent No. 6262245
: GENERAL INFORMATION:
: APPLICANT: Xu, Jiangchun
: TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE CANCER AND METHODS
: NUMBER OF SEQUENCES: 224
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: SEED and BERRY LLP
: STREET: 6300 Columbia Center, 701 Fifth Avenue
: CITY: Seattle
: STATE: WA
: COUNTRY: USA
: ZIP: 98104
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
```

```

: APPLICATION NUMBER: US/09/030,607
: FILING DATE: 25-FEB-1998
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: Haki, David J.
: REGISTRATION NUMBER: 31,392
: REFERENCE/DOCKET NUMBER: 210121.427C3
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (206) 622-4900
: TELEFAX: (206) 682-6031
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 816 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cdna
:
US-09-030-607-2

Query Match 5.0% Score 68: DB 4: Length 816:
Best Local Similarity 100.0% Pred No. 8,6e-18:
Matches 68: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

QY 1299 gccaccgcggtagagctccagcttbtccctctttagtagaggttaattacgcgccttggc 1358
      |||||||
DB 424 GCCACCGCGGTGGAGCTCCAGCTTTTGTTCCTTTAGTAGGGTTAAITTCGCGCTTGGC 483

QY 1359 gtaatcat 1366
      |||||||
DB 484 GTAATCAT 491

RESULT 10
US-08-713-000-4/c
: Sequence 4, Application US/08713000
: Patent No. 5850020
: GENERAL INFORMATION:
: APPLICANT: Bloksberg, Leonard N.
: APPLICANT: Hayukkala, Ilkka
: APPLICANT: Grierson, Alastair
: TITLE OF INVENTION: MATERIALS AND METHODS FOR THE
: TITLE OF INVENTION: MODIFICATION OF PLANT LIGNIN CONTENT
: NUMBER OF SEQUENCES: 15
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Speckman picard PLLC
: STREET: 2601 Elliott Avenue, Suite 4185
: CITY: Seattle
: STATE: WA
: COUNTRY: USA
: ZIP: 98121
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/713,000
: FILING DATE:
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Sleath, Janet
: REGISTRATION NUMBER: 37,007
: REFERENCE/DOCKET NUMBER: 11000.1003
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 206-269-0565
: TELEFAX: 206-269-0563
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 949 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear

```

## RESULT 12

US-09-211-710-4/C  
: Sequence 4, Application US/09211710A  
: Patent No. 6204434  
: GENERAL INFORMATION:  
: APPLICANT: Bloksberg, Leonard N.  
: APPLICANT: Havukkala, Ilkka  
: APPLICANT: Grierson, Alastair  
: TITLE OF INVENTION: Materials and Methods for the  
: TITLE OF INVENTION: Modification of Plant Lignin Content  
: FILE REFERENCE: 11090.1003c3  
: CURRENT APPLICATION NUMBER: US/09/211,710A  
: CURRENT FILING DATE: 1998-12-14  
: NUMBER OF SEQ ID NOS: 15  
: SOFTWARE: FastSeq for Windows Version 3.0  
: SEQ ID NO 4  
: LENGTH: 949  
: TYPE: DNA  
: ORGANISM: Pinus radiata  
: FEATURE:  
: NAME/KEY: unsure  
: LOCATION: (1)...(949)  
: OTHER INFORMATION: n at all occurrences indicates unsure

US-09-211-710-4

Query Match 5.0%; Score 68; DB 4; Length 949;  
Best Local Similarity 100.0%; Pred. No. 8.4e-18;  
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1299 gccaccggtgagctccagctttgttcccttagtgagggtaattgcgcgttggc 1358  
|||||  
Db 248 gccaccggtgagctccagctttgttcccttagtgagggtaattgcgcgttggc 189

Oy 1359 gtaatcat 1365  
|||||  
Db 188 gtaatcat 181

## RESULT 13

US-08-446-935-6  
: Sequence 6, Application US/08446935  
: Patent No. 6187991  
: GENERAL INFORMATION:  
: APPLICANT: Soeller, Walter C.  
: APPLICANT: Carty, Maynard D.  
: APPLICANT: Kreutter, David K.  
: TITLE OF INVENTION: TRANSGENIC ANIMAL MODELS FOR TYPE II  
: TITLE OF INVENTION: DIABETES MELLITUS  
: NUMBER OF SEQUENCES: 15  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: Pfizer Inc.  
: STREET: 235 East 42nd Street, 20th Floor  
: CITY: New York  
: STATE: New York  
: COUNTRY: U.S.A.  
: ZIP: 10017-5755  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: Floppy disk  
: COMPUTER: IBM PC compatible  
: OPERATING SYSTEM: PC-DOS/MS-DOS  
: SOFTWARE: PatentIn Release #1.0, Version #1.25  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/08/446,935  
: FILING DATE:  
: CLASSIFICATION: 800  
: ATTORNEY/AGENT INFORMATION:  
: NAME: Sheyka, Robert F.  
: REGISTRATION NUMBER: 31,304  
: REFERENCE/DOCKET NUMBER: PC8153  
: TELECOMMUNICATION INFORMATION:

TELEPHONE: (212)573-1189  
TELEFAX: (212)573-1939  
TELEX: N/A  
: INFORMATION FOR SEQ ID NO: 6:  
: SEQUENCE CHARACTERISTICS:  
: LENGTH: 2961 base pairs  
: TYPE: nucleic acid  
: STRANDEDNESS: double  
: TOPOLOGY: circular  
: MOLECULE TYPE: DNA (genomic)  
US-08-446-935-6

Query Match 5.0%; Score 68; DB 4; Length 2961;  
Best Local Similarity 100.0%; Pred. No. 7.2e-18;  
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1299 gccaccggtgagctccagctttgttcccttagtgagggtaattgcgcgttggc 1358  
|||||  
Db 743 gccaccggtgagctccagctttgttcccttagtgagggtaattgcgcgttggc 802

Oy 1359 gtaatcat 1365  
|||||  
Db 803 gtaatcat 810

## RESULT 14

US-08-992-334-1  
: Sequence 1, Application US/08992334  
: Patent No. 5919678  
: GENERAL INFORMATION:  
: APPLICANT: Gruss, Alexandra  
: APPLICANT: Maguin, Emmanuelle  
: TITLE OF INVENTION: METHODS FOR USING A TEMPERATURE-SENSITIVE  
: TITLE OF INVENTION: PLASMID  
: NUMBER OF SEQUENCES: 3  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: Christie Parker & Hale, LLP  
: STREET: 350 West Colorado Boulevard, Suite 500  
: CITY: Pasadena  
: STATE: California  
: COUNTRY: United States  
: ZIP: 91105  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: Floppy disk  
: COMPUTER: IBM PC compatible  
: OPERATING SYSTEM: PC-DOS/MS-DOS  
: SOFTWARE: PatentIn Release #1.0, Version #1.30  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/08/992,334  
: FILING DATE: 17-DEC-1997  
: CLASSIFICATION: 435  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: US 08/302,752  
: FILING DATE: 24-DEC-1994  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: PCT/FR93/00248  
: FILING DATE: 12-MAR-1993  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: FR 992/03034  
: FILING DATE: 13-MAR-1992  
: ATTORNEY/AGENT INFORMATION:  
: NAME: Prout, D. Bruce  
: REGISTRATION NUMBER: 20958  
: REFERENCE/DOCKET NUMBER: C93:31779  
: TELECOMMUNICATION INFORMATION:  
: TELEPHONE: (526) 795-9900  
: TELEFAX: (626) 577-8800  
: INFORMATION FOR SEQ ID NO: 1:  
: SEQUENCE CHARACTERISTICS:  
: LENGTH: 3792 base pairs  
: TYPE: nucleic acid  
: STRANDEDNESS: both

Tue Apr 30 14:18:55 2002

```

; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: YES
; IMMEDIATE SOURCE:
; CLONE: pg+host4
US-08-992-334-1

Query Match          5.0%; Score 68; DB 2; Length 3792;
Best Local Similarity 100.0%; Pred. No. 7e-18;
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1299 gccaccgcggtggagctccagcttttgccttagtgagggttaattgcgccttggc 1358
      |||||||
DB 3558 GCCACCGCGGTGGAGCTCCAGCTTTTGTTCCTTAGTGAGGGTTAATGCGCGCTTGGC 3517

QY 1359 gtaatacat 1366
      |||||||
DB 3618 GTAATCAT 3625

RESULT 15
US-08-302-752-1
; Sequence 1, Application US/08302752
; Patent No. 6025190
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: THERMOSENSIBLE PLASMID
; NUMBER OF SEQUENCES: 3
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/302,752
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 9203034
; FILING DATE: 13-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO FR/93/00248
; FILING DATE: 12-MAR-1993
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3792 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-302-752-1

Query Match          5.0%; Score 68; DB 3; Length 3792;
Best Local Similarity 100.0%; Pred. No. 7e-18;
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1299 gccaccgcggtggagctccagcttttgccttagtgagggttaattgcgccttggc 1358
      |||||||
DB 3558 GCCACCGCGGTGGAGCTCCAGCTTTTGTTCCTTAGTGAGGGTTAATGCGCGCTTGGC 3517

QY 1359 gtaatacat 1366
      |||||||
DB 3618 GTAATCAT 3625

RESULT 16
US-08-651-472-62/c
; Sequence 62, Application US/08651472
; Patent No. 5103244
; GENERAL INFORMATION:
; APPLICANT: DORNER, Friedrich
; APPLICANT: SCHEIFLINGER, Michael
; APPLICANT: PFLEIDERER, Michael
; TITLE OF INVENTION: DIRECT MOLECULAR CLONING OF CHIMERIC VIRUSES CONTAINING HUMAN IMMUNODEFICIENCY VIRUS TYPE I ANTIGENS (HIV-1) ANTIGENS
; NUMBER OF SEQUENCES: 95
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/651,472
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/914,738
; FILING DATE: 20-JUL-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/750,080
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/166/1MMU
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 62:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4145 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Other nucleic acid:
; DESCRIPTION: Synthetic DNA oligonucleotide
; IMMEDIATE SOURCE:
; CLONE: pS2qpt-S4
US-08-651-472-62

Query Match          5.0%; Score 68; DB 3; Length 4145;
Best Local Similarity 100.0%; Pred. No. 6.9e-18;
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1299 gccaccgcggtggagctccagcttttgccttagtgagggttaattgcgccttggc 1358
      |||||||
DB 2219 GCCACCGCGGTGGAGCTCCAGCTTTTGTTCCTTAGTGAGGGTTAATGCGCGCTTGGC 2160

QY 1359 gtaatacat 1366
      |||||||
DB 2159 GTAATCAT 2152

RESULT 17
US-08-358-928-62/C
; Sequence 62, Application US/08358928
; Patent No. 6265183
; GENERAL INFORMATION:
; APPLICANT: DORNER, Friedrich
; APPLICANT: SCHEIFLINGER, Michael
; APPLICANT: PFLEIDERER, Michael
```





```

SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/992.334
FILING DATE: 17-DEC-1997
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/302,752
FILING DATE: 24-DEC-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/FR93/00248
FILING DATE: 12-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR F92/03034
FILING DATE: 13-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Prout, D. Bruce
REGISTRATION NUMBER: 20958
REFERENCE/DOCKET NUMBER: C93:31779
TELECOMMUNICATION INFORMATION:
TELEPHONE: (626) 795-9900
TELEFAX: (626) 577-8800
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 5234 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
US-08-992-334-2

```

Query Match 5.0%; Score 68; DB 2; Length 5234;  
Best Local Similarity 100.0%; Pred. No. 6.7e-18;  
Matches 68; Conservative 0; Mismatches 0; Indels

**Qy.**           1299 gccaccgcggtagagctccagcttctgttcacctttagtgaagggttaattgcgcgcttgcc 1358  
             | | | | | | | | | | | | | | | |  
**b**             5000 ccccacccccctcgcactccacgttttgttcccttagtgagggttanttgccgcgcttggc 5059

```
Query Match      5.0%; Score 58; DB 4; Length 4277;
Best Local Similarity 100.0%; Pred. No. 6.9e-18;
Matches 58; Conservating 0; Mismatches 0; Indels 0; Gaps 0;
```

Oy	1299	gcacgcgggtgagactccaagtcttttggctcccttaattagtagagggtlaattatcgcgagccttgcc	1358
Db	2219	GCACGCGGTGGAGCTCCAGCTTTTGTTCCTTTAGTAGAGGTTAATTGCGCGCTTTGCC	2160
Oy	1359	gtaatcat 1366 	
Db	2159	GTAATCAT 2152 	

```

RESULT 20
US-08-992-334-2
: Sequence 2, Application US/08992334
: Patent No. 5919678
: GENERAL INFORMATION:
: APPLICANT: Gruss, Alexandra
: APPLICANT: Maquin, Emmanuelle
: TITLE OF INVENTION: METHODS FOR USING A TEMPERATURE-SENSITIVE
: TITLE OF INVENTION: PLASMID
: NUMBER OF SEQUENCES: 3
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Christie Parker & Hale, LLP
: STREET: 350 West Colorado Boulevard, Suite 500
: CITY: Pasadena
: STATE: California
: COUNTRY: United States
: ZIP: 91105
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS

```

```

RESULT      21
US-08-302-752-2
: Sequence 2, Application US/08302752
: Patent No. 6025190
: GENERAL INFORMATION:
: APPLICANT:
: TITLE OF INVENTION: THERMOSENSIBLE PLASMIID
: NUMBER OF SEQUENCES: 3
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0. Version
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/302.752
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: FR 9203034
: FILING DATE: 13-MAR-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: WO FR/93/00248
: FILING DATE: 12-MAR-1993
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 5234 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear

```

Tue Apr 30 14:18:55 2002

MOLECULE TYPE: DNA (genomic)  
US-08-302-752-2

Query Match 5.0%; Score 68; DB 3; Length 5234;

Best Local Similarity 100.0%; Pred. No. 6.7e-18;

Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1299 gccaccggtgagctccagcttttgccttttagtgagggttaattgcgcgttggc 1358

|||||

Db 5000 GCCACCGCGTGGAGCTCCAGCTTTTGTTCCTTTAGTGAAGGTTAAATGGCGCTTGGC 5059

Qy 1359 gtaatcat 1366

|||||

Db 5060 GTAATCAT 5067

RESULT 22

US-08-446-935-1

Sequence 1, Application US/08446935

Patent No. 6187991

GENERAL INFORMATION:

APPLICANT: Soeller, Walter C.

APPLICANT: Carthy, Maynard D.

APPLICANT: Kreutter, David K.

TITLE OF INVENTION: TRANSGENIC ANIMAL MODELS FOR TYPE II

TITLE OF INVENTION: DIABETES MELLITUS

NUMBER OF SEQUENCES: 15

CORRESPONDENCE ADDRESS:

ADDRESSEE: Pfizer Inc.

STREET: 235 East 42nd Street, 20th Floor

CITY: New York

STATE: New York

COUNTRY: U.S.A.

ZIP: 10017-5755

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, version #1.25

CURRENT APPLICATION NUMBER: US/08/446,935

FILING DATE:

CLASSIFICATION: 800

ATTORNEY/AGENT INFORMATION:

NAME: Shevka, Robert F.

REGISTRATION NUMBER: 31,304

REFERENCE/DOCKET NUMBER: PC8153

TELEPHONE: (212)573-1189

TELEFAX: (212)573-1939

TELEX: N/A

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 5356 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: circular

MOLECULE TYPE: DNA (genomic)

US-08-446-935-1

Query Match 5.0%; Score 68; DB 4; Length 5356;

Best Local Similarity 100.0%; Pred. No. 6.7e-18;

Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1299 gccaccggtgagctccagcttttgccttttagtgagggttaattgcgcgttggc 1358

|||||

Db 3138 GCCACCGCGTGGAGCTCCAGCTTTTGTTCCTTTAGTGAAGGTTAAATGGCGCTTGGC 197

Qy 1359 gtaatcat 1366

|||||

Db 3198 GTAATCAT 3205

RESULT 23

US-08-651-472-72/c

Sequence 72, Application US/08651472

Patent No. 6103244

GENERAL INFORMATION:

APPLICANT: DORNER, Friedrich

APPLICANT: SCHEIFLINGER, Friedrich

APPLICANT: FALKNER, Falko Gunter

APPLICANT: PFLEIDERER, Michael

TITLE OF INVENTION: DIRECT MOLECULAR CLONING OF CHIMERIC

TITLE OF INVENTION: VIRUSES CONTAINING HUMAN IMMUNODEFICIENCY VIRUS TYPE 1

TITLE OF INVENTION: (HIV-1) ANTIGENS

NUMBER OF SEQUENCES: 95

CORRESPONDENCE ADDRESS:

ADDRESSEE: Foley &amp; Lardner

STREET: 3000 K Street, N.W., Suite 500

CITY: Washington

STATE: D.C.

COUNTRY: USA

ZIP: 20007-5109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/651,472

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/914,738

FILING DATE: 20-JUL-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/750,080

FILING DATE: 26-AUG-1991

ATTORNEY/AGENT INFORMATION:

NAME: BENT, Stephen A.

REGISTRATION NUMBER: 29,768

REFERENCE/DOCKET NUMBER: 30472/166/IMMU

TELEPHONE: (202)672-5300

TELEFAX: (202)672-5399

TELEX: 904136

INFORMATION FOR SEQ ID NO: 72:

SEQUENCE CHARACTERISTICS:

LENGTH: 5532 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: Other nucleic acid;

DESCRIPTION: Synthetic DNA oligonucleotide

IMMEDIATE SOURCE:

CLONE: pAN2gptA-FIX

US-08-651-472-72

Query Match 5.0%; Score 68; DB 3; Length 5532;

Best Local Similarity 100.0%; Pred. No. 6.7e-18;

Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1299 gccaccggtgagctccagcttttgccttttagtgagggttaattgcgcgttggc 1358

|||||

Db 2219 GCCACCGCGTGGAGCTCCAGCTTTTGTTCCTTTAGTGAAGGTTAAATGGCGCTTGGC 2160

Qy 1359 gtaatcat 1366

|||||

Db 2159 GTAATCAT 2152

RESULT 24

US-08-358-928-72/c

```
; Sequence 72, Application US/08358928
; Patent No. 6265183
; GENERAL INFORMATION:
; APPLICANT: DORNER, Friedrich
; APPLICANT: SCHEIFLINGER, Friedrich
; APPLICANT: FALKNER, Falko Gunter
; APPLICANT: PFLEIDERER, Michael
; TITLE OF INVENTION: DIRECT MOLECULAR CLONING OF CHIMERIC
; TITLE OF INVENTION: VIRUSES CONTAINING HUMAN IMMUNODEFICIENCY VIRUS TYPE 1
; TITLE OF INVENTION: (HIV-1) ANTIGENS
; NUMBER OF SEQUENCES: 95
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/358,928
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/914,738
; FILING DATE: 20-JUL-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/750,080
; FILING DATE: 26-AUG-1991
; NAME: BENT, Stephen A.
; ATTORNEY/AGENT INFORMATION:
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/166/IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 72:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5532 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Other nucleic acid;
; DESCRIPTION: Synthetic DNA oligonucleotide
; IMMEDIATE SOURCE:
; CLONE: pN29pta-Fix
; US-08-358-928-72

Query Match 5.0%; Score 68; DB 4; Length 5532;
Best Local Similarity 100.0%; Pred. No. 6.7e-18;
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1299 gccaccgcggtgagctccagcttttttccctttagtgagggtaattgcgcgttggc 1358
Db 2219 gccaccgcggtgagctccagcttttttccctttagtgagggtaattgcgcgttggc 2160

QY 1359 gtaatcat 1365
Db 2159 gtaatcat 2152

RESULT 25
US-08-992-334-3
; Sequence 3, Application US/08992334
; Patent No. 5919678
; GENERAL INFORMATION:
; APPLICANT: Gruss, Alexandra
```

```
; APPLICANT: Maguin, Emmanuelle
; TITLE OF INVENTION: METHODS FOR USING A TEMPERATURE-SENSITIVE
; TITLE OF INVENTION: PLASMID
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Christie Parker & Hale, LLP
; STREET: 350 West Colorado Boulevard, Suite 500
; CITY: Pasadena
; STATE: California
; COUNTRY: United States
; ZIP: 91105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/992,334
; FILING DATE: 17-DEC-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,752
; FILING DATE: 24-DEC-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/FR93/00248
; FILING DATE: 12-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 92/03034
; FILING DATE: 13-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Prout, D. Bruce
; REGISTRATION NUMBER: 20958
; REFERENCE/DOCKET NUMBER: C93:31779
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (626) 795-9900
; TELEFAX: (626) 577-8800
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6722 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; US-08-992-334-3

Query Match 5.0%; Score 68; DB 2; Length 6722;
Best Local Similarity 100.0%; Pred. No. 6.5e-18;
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1299 gccaccgcggtgagctccagcttttttccctttagtgagggtaattgcgcgttggc 1358
Db 5488 gccaccgcggtgagctccagcttttttccctttagtgagggtaattgcgcgttggc 6547

QY 1359 gtaatcat 1365
Db 5488 gtaatcat 6555

RESULT 26
US-08-302-752-3
; Sequence 3, Application US/08302752
; Patent No. 6025190
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: THERMOSENSIBLE PLASMID
; NUMBER OF SEQUENCES: 3
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
```



LENGTH: 6811 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: Other nucleic acid:  
DESCRIPTION: Synthetic DNA oligonucleotide  
IMMEDIATE SOURCE:  
CLONE: pN2-gp1a Prots  
US-08-358-928-67

Query Match 5.0%; Score 68; DB 4; Length 6811;  
Best Local Similarity 100.0%; Pred. No. 6.5e-18;  
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1299 gccacccggtgagctccagcttttgccttttagtgaggttaattgcgcgttggc 1358  
|||||  
Db 2219 gccacccggtgagctccagcttttgccttttagtgaggttaattgcgcgttggc 2160

Oy 1359 gtaatacat 1366  
|||||  
Db 2159 gtaatacat 2152

## RESULT 29

US-08-651-472-69/c  
Sequence 69, Application US/08651472  
Patent No. 6103244

## GENERAL INFORMATION:

APPLICANT: DORNER, Friedrich  
APPLICANT: SCHEIFLINGER, Friedrich  
APPLICANT: FALKNER, Falko Gunter  
APPLICANT: PELEIDERER, Michael  
TITLE OF INVENTION: DIRECT MOLECULAR CLONING OF CHIMERIC  
TITLE OF INVENTION: VIRUSES CONTAINING HUMAN IMMUNODEFICIENCY VIRUS TYPE 1  
TITLE OF INVENTION: (HIV-1) ANTIGENS  
NUMBER OF SEQUENCES: 95  
CORRESPONDENCE ADDRESS:

ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W., Suite 500  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20007-5109

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/651,472  
FILING DATE:

## CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/914,738

FILING DATE: 20-JUL-1992

## PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/750,080

FILING DATE: 26-AUG-1991

## ATTORNEY/AGENT INFORMATION:

NAME: BENT, Stephen A.

REGISTRATION NUMBER: 29,768

REFERENCE/DOCKET NUMBER: 30472/166/IMMU

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202)672-5300

TELEFAX: (202)672-5399

TELEX: 904135

INFORMATION FOR SEQ ID NO: 59:

## SEQUENCE CHARACTERISTICS:

LENGTH: 6926 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: Other nucleic acid:  
DESCRIPTION: Synthetic DNA oligonucleotide  
IMMEDIATE SOURCE:  
CLONE: pP2-gp160MN  
US-08-651-472-69

Query Match 5.0%; Score 68; DB 3; Length 6926;  
Best Local Similarity 100.0%; Pred. No. 6.5e-18;  
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1299 gccacccggtgagctccagcttttgccttttagtgaggttaattgcgcgttggc 1358  
|||||  
Db 2219 gccacccggtgagctccagcttttgccttttagtgaggttaattgcgcgttggc 2160

Oy 1359 gtaatacat 1366  
|||||  
Db 2159 gtaatacat 2152

## RESULT 30

US-08-358-928-69/c  
Sequence 69, Application US/08358928  
Patent No. 6265183

## GENERAL INFORMATION:

APPLICANT: DORNER, Friedrich  
APPLICANT: SCHEIFLINGER, Friedrich  
APPLICANT: FALKNER, Falko Gunter  
APPLICANT: PELEIDERER, Michael  
TITLE OF INVENTION: DIRECT MOLECULAR CLONING OF CHIMERIC  
TITLE OF INVENTION: VIRUSES CONTAINING HUMAN IMMUNODEFICIENCY VIRUS TYPE 1  
TITLE OF INVENTION: (HIV-1) ANTIGENS  
NUMBER OF SEQUENCES: 95  
CORRESPONDENCE ADDRESS:

ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W., Suite 500  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20007-5109

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/358,928  
FILING DATE:

## CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/914,738

FILING DATE: 20-JUL-1992

## PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/750,080

FILING DATE: 26-AUG-1991

## ATTORNEY/AGENT INFORMATION:

NAME: BENT, Stephen A.

REGISTRATION NUMBER: 29,768

REFERENCE/DOCKET NUMBER: 30472/166/IMMU

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202)672-5300

TELEFAX: (202)672-5399

TELEX: 904136

INFORMATION FOR SEQ ID NO: 69:

## SEQUENCE CHARACTERISTICS:

LENGTH: 6926 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: Other nucleic acid:

DESCRIPTION: Synthetic DNA oligonucleotide

IMMEDIATE SOURCE:

CLONE: pP2-gp160MN



; Sequence 135, Application US/08998416  
; Patent No. 6239264  
; GENERAL INFORMATION:  
; APPLICANT: Philippsen, Peter  
; APPLICANT: Pohlmann, Rainer  
; APPLICANT: Steiner, Sabine  
; APPLICANT: Mohr, Christine  
; APPLICANT: Wendland, Jurgon  
; APPLICANT: Knechtle, Philipp  
; APPLICANT: Reibschung, Corinne  
; TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHEYA GOSSYPIL  
; TITLE OF INVENTION: AND USES THEREOF  
; NUMBER OF SEQUENCES: 1152  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: NO. 6239264artis Corporation  
; STREET: 3054 Cornwalis Road  
; CITY: Research Triangle Park  
; STATE: No. 6239264th Carolina  
; COUNTRY: USA  
; ZIP: 27709  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/998,416  
; FILING DATE: 24-DEC-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: CH 0016/97  
; FILING DATE: 31-DEC-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Meigs, J. Timothy  
; REGISTRATION NUMBER: 38,241  
; REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 919-541-8587  
; TELEFAX: 919-541-8689  
; INFORMATION FOR SEQ ID NO: 135:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 713 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; ORIGINAL SOURCE:  
; ORGANISM: PAG1051RP  
; US-08-998-416-135

Query Match 4.5%: Score 62; DB 4; Length 713;  
Best Local Similarity 100.0%; Pred. No. 1.8e-15;  
Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 ccagtgcgcgcgtaatacagcactcactataggcggaattgggtaccgggccccccctcg 60  
|||||  
Db 246 CCAGTGAGCGCGCGTAATACGACTACACTATAGGGCGGAATTGGTACCGGGCCCCCTCG 187  
Qy 61 ag 62  
||  
Db 186 AG 185

RESULT 34  
US-08-452-267-3/C  
; Sequence 3, Application US/08452267  
; Patent No. 5801027  
; GENERAL INFORMATION:  
; APPLICANT: Bennett, Malcolm  
; APPLICANT: May, Sean  
; APPLICANT: Ramsay, Nichola  
; TITLE OF INVENTION: Control of Genes in Transgenic Plants

; NUMBER OF SEQUENCES: 3  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Barnes & Thornburg  
; STREET: 11 South Meridian  
; CITY: Indianapolis  
; STATE: IN  
; COUNTRY: USA  
; ZIP: 46204  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/452,267  
; FILING DATE: 26-MAY-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Breen, John P.  
; REGISTRATION NUMBER: 38,833  
; REFERENCE/DOCKET NUMBER: 6653-25744  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (317) 231-7745  
; TELEFAX: (317) 231-7433  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 5534 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: circular  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; IMMEDIATE SOURCE:  
; CLONE: PUMIGIT  
; US-08-452-267-3

Query Match 4.5%: Score 62; DB 1; Length 5534;  
Best Local Similarity 100.0%; Pred. No. 1.4e-15;  
Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 ccagtgcgcgcgtaatacagcactcactataggcggaattgggtaccgggccccccctcg 60  
|||||  
Db 2717 CCAGTGAGCGCGCGTAATACGACTACACTATAGGGCGGAATTGGTACCGGGCCCCCTCG 2658  
Qy 61 ag 62  
||  
Db 2657 AG 2656

RESULT 35  
US-09-123-644-3/C  
; Sequence 3, Application US/09123644  
; Patent No. 5127606  
; GENERAL INFORMATION:  
; APPLICANT: Bennett, Malcolm  
; APPLICANT: May, Sean  
; APPLICANT: Ramsay, Nichola  
; TITLE OF INVENTION: Method of Using Transactivation Proteins to  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Barnes & Thornburg  
; STREET: 11 South Meridian  
; CITY: Indianapolis  
; STATE: IN  
; COUNTRY: USA  
; ZIP: 46204  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25

```

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/123,644
FILING DATE: 28-JUL-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Breen, John P.
REGISTRATION NUMBER: 38,833
REFERENCE/DOCKET NUMBER: 6053-60788
TELECOMMUNICATION INFORMATION:
TELEPHONE: (317) 231-7745
TELEFAX: (317) 231-7433
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 5534 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
IMMEDIATE SOURCE:
CLONE: pUMIGIT
US-09-123-644-3

```

```

Query Match      4.5% Score 62 DB 3 Length 5534;
Best Local Similarity 100.0% Pred. No. 1.4e-15;
Matches 62: Conservative 0 Mismatches 0 Indels 0 Gaps 0;

QY 1 ccagtgagcgcgcgtaatacgaactcaactatagggcgaattgggtatcgggccccctccq 50
|||||
Db 2717 CCATGAGGCGCGTAATACCACTCACTATAGGGCGAAITGGGTATCGGGCCCCCCCCICG 2658
|||||

QY 61 ag 62
|||
Db 2657 AG 2656

```

RESULT 36  
US-09-128-314-4/c  
Sequence 4, Application US/09128314  
Patent No. 6183121  
GENERAL INFORMATION:  
APPLICANT: Kim, Joseph L.  
APPLICANT: Morgenstern, Kurt A  
APPLICANT: Caron, Paul R  
APPLICANT: Lin, Chao  
APPLICANT: Vertex Pharmaceuticals Inc.  
TITLE OF INVENTION: CRYSTAL STRUCTURE OF THE HCV NS3 HELICASE DOMAIN  
FILE REFERENCE: Sequence listing for VPI/97-101  
Patent No. 6183121  
CURRENT APPLICATION NUMBER: US/09/128,314  
CURRENT FILING DATE: 1998-08-03  
EARLIER APPLICATION NUMBER: 60/055,772  
EARLIER FILING DATE: 1997-08-13  
NUMBER OF SEQ ID NOS: 4  
SOFTWARE: PatentIn ver. 2.0  
SEQ ID NO 4  
LENGTH: 7659  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: Genetically engineered plasmid containing helicase domain of HCV NS3  
OTHER INFORMATION: HCV NS3  
US-09-128-314-4

```

Query Match      4.5%; Score 62; DB 4; Length 7659;
Best Local Similarity 100.0%; Pred. No. 1.3e-15;
Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Ov 1 ccagtgaaacgcgcgaataacacactcactatagaagcgaattggctcaggccccctcgg 60

```

```

Db      1900  CCAGTCGAGCGCGTAATACGACTCACTATAGGCGCAATTGGGTACGGGCCCCCCTCG 1841
      |||||||
Qf:      61  ag  62
      ||
Db      1840  AG 1839
      ||
RESULT  37
US-09-128-314-3/c
: Sequence 3, Application US/09128314
: Patent No. 6183121
: GENERAL INFORMATION:
: APPLICANT: Kim, Joseph L
: APPLICANT: Morgenstern, Kurt A
: APPLICANT: Caron, Paul R
: APPLICANT: Lin, Chao
: APPLICANT: Vertex Pharmaceuticals Inc.
: TITLE OF INVENTION: CRYSTAL STRUCTURE OF THE HCV NS3 HELICASE DOMAIN
: FILE REFERENCE: Sequence listing for VPI/97-101
: Patent No. 6183121
: CURRENT APPLICATION NUMBER: US/09/128,314
: CURRENT FILING DATE: 1998-08-03
: EARLIER APPLICATION NUMBER: 60/055,772
: EARLIER FILING DATE: 1997-08-13
: NUMBER OF SEQ ID NOS: 4
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 3
: LENGTH: 8157
: TYPE: DNA
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Description of Artificial Sequence: Genetically
: OTHER INFORMATION: engineered plasmid containing full-length HCV NS3
: OTHER INFORMATION: coding sequence
US-09-128-314-3

```

```

Query Match          4.5%: Score 62: DB 4: Length 8157;
Best Local Similarity 100.0%: Pred. No. 1.3e-15;
Matches 52ex: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

Q: 1 ccagtgagcgcgcgtaatacgaactcaactataagcgcaattgggtaccggcgccccctcg 60
  |||||
DB 2398 FCAGTGAAGCGCGTAAACGACTCACATAGGGCGAAATGGGTACCGGGCCCCCCCCICG 2339

Q: 61 ag 62
  ||
DB 2398 AG 2337

```

```

RESULT 38
US-08-095-947-2/c
: Sequence 2, Application US/08096947
: Patent No. 5506137
: GENERAL INFORMATION:
: APPLICANT: Eric J. Mathur
: APPLICANT: Edward J. Marsh
: APPLICANT: Warren E. Schoettlin
: TITLE OF INVENTION: Purified thermostable
: TITLE OF INVENTION: Pyrococcus furiosus DNA
: TITLE OF INVENTION: Ligase
: NUMBER OF SEQUENCES: 17
: CORRESPONDENCE ADDRESSES:
: ADDRESSEE: Limbach & Limbach
: STREET: 2001 Ferry Building
: CITY: San Francisco
: STATE: CA
: COUNTRY: USA
: ZIP: 94111
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette, 3+ inch, 1.4 Mb storage
: COMPUTER: IBM PC Compatible

```



OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WordPerfect 5.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/096,947  
FILING DATE: 19930722  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/919,140  
FILING DATE: July 23, 1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Albert P. Halluin  
REGISTRATION NUMBER: 25,227  
REFERENCE/DOCKET NUMBER: STRG 20081 USA  
TELEPHONE: (415) 433-4150  
TELEFAX: (415) 433-8715  
TELEX: 278356  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2363 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: genomic DNA  
ANTI-SENSE: no  
ORIGINAL SOURCE:  
ORGANISM: Pyrococcus Furiosus  
INDIVIDUAL ISOLATE: DSM #3638  
CELL TYPE: unicellular organism  
IMMEDIATE SOURCE:  
LIBRARY: Pyrococcus Furiosus Genomic DNA  
CLONE: pEMI  
US-08-096-947-2

Query Match 4.5%: Score 61: DB 1: Length 2363;  
Best Local Similarity 100.0%: Pred. No. 3.7e-15;  
Matches 61: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

QY 1299 gccaccgcggtgagctccagctttgtcccttagtgagggttaattgcgcgttggc 1358  
|||||  
DB 141 GCCACCGCGGTGAGCTCCAGCTTTGTTCCTTTAGTGAGGGTTAATTCGCCGCTTGGC 82

QY 1359 g 1359  
DB 81 G 81

RESULT 39  
US-07-919-140B-2/c  
Sequence 2, Application US/07919140B  
Patent No. 5700672  
GENERAL INFORMATION:  
APPLICANT: Eric J. Mathur  
APPLICANT: Edward J. Marsh  
APPLICANT: Warren E. Schoettlin  
TITLE OF INVENTION: Purified Thermostable Pyrococcus  
NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds  
STREET: 2730 Sand Hill Road  
CITY: Menlo Park  
STATE: California  
COUNTRY: USA  
ZIP: 94025  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy Disk  
COMPUTER: IBM PC Compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WordPerfect 5.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/919,140B

FILING DATE: July 23, 1992  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Albert P. Halluin  
REGISTRATION NUMBER: 25,227  
REFERENCE/DOCKET NUMBER: 8142-013  
TELEPHONE: 415-854-3660  
TELEFAX: 415-854-3694  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2363 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: genomic DNA  
ANTI-SENSE: no  
ORIGINAL SOURCE:  
ORGANISM: Pyrococcus Furiosus  
INDIVIDUAL ISOLATE: DSM #3638  
CELL TYPE: unicellular organism  
IMMEDIATE SOURCE:  
LIBRARY: PTO.MBPpyrococcus Furiosus Genomic DNA  
CLONE: pEMI  
US-07-919-140B-2

Query Match 4.5%: Score 61: DB 1: Length 2363;  
Best Local Similarity 100.0%: Pred. No. 3.7e-15;  
Matches 61: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

QY 1299 gccaccgcggtgagctccagctttgtcccttagtgagggttaattgcgcgttggc 1358  
|||||  
DB 141 GCCACCGCGGTGAGCTCCAGCTTTGTTCCTTTAGTGAGGGTTAATTCGCCGCTTGGC 82

QY 1359 g 1359  
DB 81 G 81

RESULT 40  
US-08-915-232-2/c  
Sequence 2, Application US/08916232  
Patent No. 6280998  
GENERAL INFORMATION:  
APPLICANT: Eric J. Mathur  
APPLICANT: Edward J. Marsh  
APPLICANT: Warren E. Schoettlin  
TITLE OF INVENTION: Purified Thermostable Pyrococcus  
NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds  
STREET: 2730 Sand Hill Road  
CITY: Menlo Park  
STATE: California  
COUNTRY: USA  
ZIP: 94025  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy Disk  
COMPUTER: IBM PC Compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WordPerfect 5.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/916,232  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/919,140  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Albert P. Halluin

TELEFAX: (415) 433-8716  
TELEX: 278356  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2363 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: genomic DNA  
ANTI-SENSE: no  
ORIGINAL SOURCE:  
ORGANISM: Pyrococcus Furiosus  
INDIVIDUAL ISOLATE: DSM #3638  
CELL TYPE: unicellular organism  
IMMEDIATE SOURCE:  
LIBRARY: Pyrococcus Furiosus Genomic DNA  
CLONE: pEM1  
PCT-US93-06939-2

Query Match 4.5% Score 61: DB 5: Length 2363:  
Best Local Similarity 100.0%: Pred. No. 3.7e-15:  
Matches 61: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

OY 1299 gccaccgcggtgagctccagcttttgcctttagtgagggttaattgcgcgttggc 1358  
DB 141 gccaccgcggtgagctccagcttttgcctttagtgagggttaattgcgcgttggc 82

OY 1359 g 1359  
DB 81 G 81

RESULT 42  
US-08-799-569-1  
Sequence 1, Application US/08799569  
Patent No. 6133244  
GENERAL INFORMATION:  
APPLICANT: Michel, Marie-Louise  
TITLE OF INVENTION: Nucleotide Vector, Composition  
TITLE OF INVENTION: Containing Such Vector, and Vaccine for Immunization  
TITLE OF INVENTION: Against Hepatitis  
NUMBER OF SEQUENCES: 1  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &  
ADDRESSEE: Dunner  
STREET: 1300 I Street, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20005-3315  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/799,569  
FILING DATE: 12-FEB-1997  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/706,337  
FILING DATE: 30-AUG-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/633,821  
FILING DATE: 22-APR-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: FR 94/00483  
FILING DATE: 27-APR-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Meyers, Kenneth J  
REGISTRATION NUMBER: 25,146

REGISTRATION NUMBER: 25,227  
REFERENCE/DOCKET NUMBER: 8142-013  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-854-3560  
TELEFAX: 415-854-3594  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2363 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: genomic DNA  
ANTI-SENSE: no  
ORIGINAL SOURCE:  
ORGANISM: Pyrococcus Furiosus  
INDIVIDUAL ISOLATE: DSM #3638  
CELL TYPE: unicellular organism  
IMMEDIATE SOURCE:  
LIBRARY: PTO.MBPyrococcus Furiosus Genomic DNA  
CLONE: pEM1  
US-08-916-232-2

Query Match 4.5% Score 61: DB 4: Length 2363:  
Best Local Similarity 100.0%: Pred. No. 3.7e-15:  
Matches 61: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

OY 1299 gccaccgcggtgagctccagcttttgcctttagtgagggttaattgcgcgttggc 1358  
DB 141 gccaccgcggtgagctccagcttttgcctttagtgagggttaattgcgcgttggc 82

OY 1359 g 1359  
DB 81 G 81

RESULT 41  
PCT-US93-06939-2/C  
Sequence 2, Application PC/TUS9306939  
GENERAL INFORMATION:  
APPLICANT: Eric J. Mathur  
APPLICANT: Edward J. Marsh  
APPLICANT: Warren E. Schoettlin  
TITLE OF INVENTION: Purified Thermostable  
TITLE OF INVENTION: Pyrococcus Furiosus DNA  
TITLE OF INVENTION: Ligase  
NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Limbach & Limbach  
STREET: 2001 Ferry Building  
CITY: San Francisco  
STATE: CA  
COUNTRY: USA  
ZIP: 94111  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3+ inch, 1.4 Mb storage  
COMPUTER: IBM PC Compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WordPerfect 5.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/06939  
FILING DATE: 19930722  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/919,140  
FILING DATE: July 23, 1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Albert P. Halluin  
REGISTRATION NUMBER: 25,227  
REFERENCE/DOCKET NUMBER: STRG 20081 USA  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 433-4150

REFERENCE/DOCKET NUMBER: 03495.0128-01000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 408-4000  
TELEFAX: (202) 408-4400  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5618 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-799-569-1

Query Match 4.0%; Score 55; DB 3; Length 5618;  
Best Local Similarity 100.0%; Pred. No. 5.8e-13;  
Matches 55; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1299 gccaccgggtgagctccagcttttattcccttttagtgaagggttaattcgacac 1353  
|||||  
Db 2803 gccaccgggtgagctccagcttttattcccttttagtgaagggttaattcgacac 2857

## RESULT 43

US-08-703-809-9/c  
Sequence 9, Application US/08703809  
Patent No. 5716808  
GENERAL INFORMATION:  
APPLICANT: Raymond, Christopher K.  
TITLE OF INVENTION: GENETIC ENGINEERING OF PICHIA METHANOLICA  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: ZymoGenetics, Inc.  
STREET: 1201 Eastlake Avenue East  
CITY: Seattle  
STATE: WA  
COUNTRY: USA  
ZIP: 98102

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Parker, Gary E  
REGISTRATION NUMBER: 31-648  
REFERENCE/DOCKET NUMBER: 96-18  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 206-442-6673  
TELEFAX: 206-442-6678  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 329 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear

Query Match 3.7%; Score 51; DB 1; Length 329;  
Best Local Similarity 100.0%; Pred. No. 3.4e-11;  
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1316 ccagcttttgccttttagtgagggtaattgcgcgttgccgtaatacat 1366  
|||||  
Db 55 CCAGCTTTTGTCCCTTTAGTGAGGGTTAATTCGCGCTTGCGGTAATCAT 5

## RESULT 44

US-08-703-808-9/c  
Sequence 9, Application US/08703808  
Patent No. 5736383  
GENERAL INFORMATION:  
APPLICANT: Raymond, Christopher K.  
TITLE OF INVENTION: PREPARATION OF PICHIA METHANOLICA AUXOTROPHIC MUTANTS  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: ZymoGenetics, Inc.  
STREET: 1201 Eastlake Avenue East  
CITY: Seattle  
STATE: WA  
COUNTRY: USA  
ZIP: 98102

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Parker, Gary E  
REGISTRATION NUMBER: 31-648  
REFERENCE/DOCKET NUMBER: 96-17  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 206-442-6673  
TELEFAX: 206-442-6678  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 329 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear

Query Match 3.7%; Score 51; DB 1; Length 329;  
Best Local Similarity 100.0%; Pred. No. 3.4e-11;  
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1316 ccagcttttgccttttagtgagggtaattgcgcgttgccgtaatacat 1366  
|||||  
Db 55 CCAGCTTTTGTCCCTTTAGTGAGGGTTAATTCGCGCTTGCGGTAATCAT 5

## RESULT 45

US-08-703-807-9/c  
Sequence 9, Application US/08703807  
Patent No. 5955349  
GENERAL INFORMATION:  
APPLICANT: Raymond, Christopher K.  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR PRODUCING HETEROLOGOUS POLYPEPTIDES IN PICHIA METHANOLICA  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: ZymoGenetics, Inc.  
STREET: 1201 Eastlake Avenue East  
CITY: Seattle  
STATE: WA  
COUNTRY: USA  
ZIP: 98102

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
FILING DATE:

CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Parker, Gary E  
REGISTRATION NUMBER: 31-548  
REFERENCE/DOCKET NUMBER: 96-16  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 206-442-6573  
TELEFAX: 206-442-6578  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 329 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-08-703-807-9

Query Match 3.7% Score 51; DB 2: Length 329:  
Best Local Similarity 100.0%; Pred. No. 3.4e-11;  
Matches 51: Conservative 0; Mismatches 0; Indels 0; Gaps 0:

Oy 1316 ccagctttgtcccttagtgagggttaattgcgcgcttgacgtaatacat 1366  
|||||  
Db 55 CCAGCTTTGTCCTTTAGTGAGGGTTAATTCGCGCTTGGCGTAATCAI 5

Search completed: April 30, 2002, 10:38:12  
Job time: 9961 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic : nucleic search, using sw model

Run on: April 29, 2002, 17:16:47 : Search time 5783.7 Seconds  
(without alignments)  
2721.794 Million cell updates/sec

Title: US-09-248-178-61  
Perfect score: 1368  
Sequence: 1 ccagtagcgcgcgaataac.....cgcccttgccgaatacatnn 1368

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 13736207.seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST:\*

- 1: em\_estba:\*
- 2: em\_esthum:\*
- 3: em\_estin:\*
- 4: em\_estmu:\*
- 5: em\_estov:\*
- 6: em\_estpl:\*
- 7: em\_estro:\*
- 8: em\_htc:\*
- 9: gb\_est1:\*
- 10: gb\_est2:\*
- 11: gb\_htc:\*
- 12: gb\_gss:\*
- 13: em\_gss\_hum:\*
- 14: em\_gss\_inv:\*
- 15: em\_gss\_pln:\*
- 16: em\_gss\_vrt:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	551.8	40.3	556	9	AI927004 w068g06.x
2	515.4	37.7	575	9	AW374095 PM4-BT054
3	499.5	36.5	550	9	AW374082 PM4-BT054
4	393.4	28.8	905	10	BE872674 501451222
5	388.8	28.4	409	9	AI085849 ox47f12.s
6	352	25.7	364	9	AI867547 wb78c04.x
7	234.4	17.1	746	12	AG030921 Pan trogl
8	126.4	9.2	661	10	BM391501 UI-R-DY0-
9	122	8.9	636	10	BM384329 UI-R-DY0-
10	119	8.7	745	11	AK003832 Mus muscu
11	110.8	8.1	369	9	AW787180 120924 MA
12	99.4	7.3	733	10	BM391251 UI-R-DY0-
13	92.4	6.8	535	9	AW177437 AU177437
14	92	6.7	658	12	AG054901 Pan trogl
15	88.9	6.5	677	9	AK720437 C01755-F
16	87.2	6.4	471	9	AW960207
17	85.2	6.2	311	12	AZ049564 GSSBru013

C 18	83	6.1	653	9	AU177453
C 19	82.2	6.0	627	10	BE414300
C 20	82.2	6.0	800	10	BE421927
C 21	81.6	6.0	253	10	BI784504
C 22	81.6	6.0	358	9	AU069050
C 23	81.6	6.0	363	9	AU069112
C 24	81.6	6.0	372	10	D80022
C 25	81.6	6.0	394	9	AU069190
C 26	81.2	5.9	385	9	AW951430
C 27	81.2	5.9	450	10	D59275
C 28	81.2	5.9	945	12	CNS04D0K
C 29	80.6	5.9	325	9	AU069730
C 30	80.6	5.9	467	9	AV746034
C 31	80.6	5.9	640	10	BE414138
C 32	80	5.8	438	10	BI945663
C 33	80	5.8	520	10	D50995
C 34	79.8	5.8	233	12	A0012365
C 35	79.6	5.8	311	9	AU068988
C 36	79.6	5.8	403	9	AW951437
C 37	79.6	5.8	421	9	AW951452
C 38	79	5.8	327	9	AU069088
C 39	79	5.8	329	10	BI118265
C 40	79	5.8	332	9	AU069089
C 41	79	5.8	476	10	BG662546
C 42	79	5.8	480	9	AW687437
C 43	79	5.8	493	9	AW954068
C 44	79	5.8	711	10	BE414359
C 45	79	5.8	732	10	BE421895

ALIGNMENTS

RESULT 1  
LOCUS AI927004 556 bp mRNA linear EST 08-NAR-2000  
DEFINITION w068g06.x1 NCI\_CGAP\_Pr22 Homo sapiens cDNA clone IMAGE:2460538 3', mRNA sequence.  
ACCESSION AI927004  
VERSION AI927004.1 GI:5662968  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 556)  
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TUMOR Tumor Gene Index  
JOURNAL National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation: M. Bento Soares, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www.bio.llnl.gov/bbrp/image/image.html  
Insert Length: 989 Std Error: 0.00  
Seq primer: -40UP from Gibco  
High quality sequence stop: 454.

FEATURES

Location/Qualifiers  
1..556  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2460538"  
/clone\_lib="NCI\_CGAP\_Pr22"  
/sex="male"  
/tissue.type="normal prostate"  
/lab\_host="DH10B"  
/note="Organ: prostate; Vector: pT7T3D-Pac (Pharmacia)

with a modified polylinker: 1st strand cDNA was prepared from normal prostate bulk tissue, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library is normalized, and was constructed by Bento Soares and M. Fatima Bonaldo.

BASE COUNT 205 a 82 c 84 g 184 t 1 others  
ORIGIN  
Query Match 40.3%; Score 551.8; DB 9; Length 556;  
Best Local Similarity 99.5%; Pred. No. 8.6e-83;  
Matches 553: Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 85 ttattatgacgaattcaggctttattattgagcaatgaaacagcctaaacttaattcc 144  
Db 1 tttattgatacaatcaggctttattattgagcaatgaaacagcctaaacttaattcc 60  
Oy 145 aagcatgtgtgttaaaagtttgcaaaagtgagatattgttcacaaacacattcaatgttt 204  
Db 61 AAGCATGTGTAGTAAAGTTTGCAAAAGTGGGATATTGTTCAAAACACATTCATGTTT 120  
Oy 205 aaacactatttattgaaacaaatattttaaattgtttcttctaaagagcccat 264  
Db 121 AAACACTATTATTGGAAGACAAATATATTAAATGTTTGTCTTAAAGCCCAT 180  
Oy 265 ttccctcaagctcaaaactttgttaatttgatattaaagcaatgaattatttgtacatc 324  
Db 181 TTCCCTCCAAAGCTAAACTTTTGTAATTGTATATTAAAGCAATGAAGTTATTGTACATT 240  
Oy 325 tagttaacaacagcaatgacagcagcagcaataaaattgcacagcagcagcagcagcagc 384  
Db 241 TAGTTAAACACAGCAATAGCACTAGGCAGCAATATAAATTTGCACAGACGTATGCAATTT 300  
Oy 385 tccagatagcattctttaaattccagttttcagcttccaaagattggttgcccaataatag 444  
Db 301 TCCAAGATAGCATCTTTAAATTCAGTTTTCAGCTTCCAAAGATTGTTGCCCAATAATAG 360  
Oy 445 acttaaacataaataagtggtcctaaacaaataagtatcgaataatgaaagagaaatg 504  
Db 361 ACTTAAACATATAATGATGGCTTAAAGAAATTAAGTATACGAAATGTAAAGAGGAATG 420  
Oy 505 taagtcactctcaactctcaataaaaggtgagagtaagatgctaaagcacaataaatgta 564  
Db 421 TAAGTCCACTCTCAATCATAAAAGGTGAGAGTAAGGATGCTAAAGCAAAATAATGTA 480  
Oy 565 ggtctttttctgtttctcgtttatcatgaatctgcttcttctgtatgctttaggatt 624  
Db 481 GGTTCTTTTTTCTGTTTCCGTTTATCATGCAATCTGCTTCTTCTTATGCTTAGGGTT 540  
Oy 625 acccatttaagttaga 640  
Db 541 ACCCAITTAAGTTAGA 556

RESULT 2  
AW374095  
LOCUS PM4-BT0548-171299-001-f08 BT0548 Homo sapiens cDNA, mRNA sequence.  
DEFINITION AW374095  
ACCESSION AW374095  
VERSION AW374095.1 GI:5678749  
KEYWORDS EST.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 576)  
REFERENCE HCGP <http://www.ludwig.org.br/OPRESIES>.  
AUTHORS The FAPESP/LICR Human Cancer Genome Project  
TITLE Unpublished (1999)  
JOURNAL Contact: Simpson A.J.G.  
COMMENT Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: [asimpson@ludwig.org.br](mailto:asimpson@ludwig.org.br)  
This sequence was derived from the FAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL  
(<http://www.ludwig.org.br/scripts/gethtml2.pl?tl=PM4&t2=PM4-BT0548-171299-001-f08&t3=1999-12-17&t4=1>)  
Seq primer: puc 18 forward  
High quality sequence stop: 575.

FEATURES  
Location/Qualifiers  
1..576  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="BT0548"  
/dev\_stage="Adult"  
/note="Organ: breast; Vector: puc18; Site: 1; SmaI: Site: 2;  
SmaI: A mini-library was made by cloning products derived  
from OPRESIES PCR (U.S. Letters Patent application No. 196  
716 - Ludwig Institute for Cancer Research) profiles  
into the pUC 18 vector. Reverse transcription of tissue  
mRNA and cDNA amplification were performed under low  
stringency conditions."

BASE COUNT 209 a 90 c 96 g 179 t 2 others  
ORIGIN

Query Match 37.7%; Score 516.4; DB 9; Length 576;  
Best Local Similarity 98.4%; Pred. No. 6.9e-77;  
Matches 553: Conservative 0; Mismatches 6; Indels 3; Gaps 3;  
Oy 180 ttttcacaaacacattcaatgttttaaacactatttatttgaagaacaaatatttaa 239  
Db 18 TGTTCAAAACACATTCATGTTTAAACACTATTATTATTTGAAGACACAAATATTGAA 77  
Oy 240 aattgtttgtcttaaaagcccatctccctcccaagctctaaacttgaatttgatctaa 299  
Db 78 AATTG-TIGTCTTAAAGGCCATTTCCCTCCAAAGTCTAAACTTTGTAATTGATATTA 136  
Oy 300 agcaataaattatttgtacaacttagttaaaacagcagaatagcactagcagaataa 359  
Db 137 AGCAATGAAGTTATTTTGTACAACTCTAGTTTAAACAGCAGATAGGACTAGGCAGAAATA 196  
Oy 360 aaaaattcacagcagcagatgcaattttccaaagatagcattctttaaattcagtttcagct 419  
Db 197 AAAATTGCACAGACGCGATGCAATTTTCCAAAGATAGCAATCTTTAAATTCAGTTTTCAGCT 256  
Oy 420 tccaaagattggttgcccaataatagcatttaaacataatagcagcttaaaaaataaagt 479  
Db 257 TCCAAAGATTGGTTGCCCATTAATAGACTTAAACATAATAATGATGGCT-AAAAAATAAGT 315  
Oy 480 atacgaaatgtaaaaaggaatgtaagtcacactctcaatctcataaaagggtgagagta 539  
Db 315 ATACGAAAAATGTAAGGAAGGAATGTAAGTCCACTCTCAATCTCATAAAGGTGAGAGTA 375  
Oy 540 aggatgtcaaaagcaaaataaattagatttctttttcttcttccgtttatcatgaact 599  
Db 376 AGGATGTAAAGCAAAATAAATGATAGGTTCTTTTTTCTATTTCCTGTTTATCATGCATC 435  
Oy 600 tgcctcttgatagccttagggttacccatttaagtttagaggttgtaagcattggtg 659  
Db 436 TGCITCTTGATATGCCCTTAGGGTTACCCATTTAAGTTAGAGGTCGTAATGCAATGGTGG 495  
Oy 650 gaatgaataatgatcaaatataacacctgtcatttctcatttcaaatgcggcgtggaact 719  
Db 495 GAATGAAAAATGATCAAAATATACACCTTGCTGTCATTTCATTTCCAAATGGC-GGCTGGAANCT 554  
Oy 720 tccaaaaaaagggttaggcata 741  
Db 555 TCCAAAAAAGGTTAGGCATGA 576

```

RESULT 3
AM374082
LOCUS PM4-BT0548-171259-001-a04 BT0548 Homo sapiens cDNA, mRNA sequence.
DEFINITION AM374082
ACCESSION AM374082
VERSION AM374082.1 GI:5878736
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 550)
HGCP http://www.ludwig.org.br/OPRESTES.
The FAPESP/LICR Human Cancer Genome Project
Unpublished (1999)
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=PM4&st2=PM4-BT0548-
171259-001-a04&st3=1999-12-17&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 13
High quality sequence stop: 549.
Features
Location/Qualifiers
1..550
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="BT0548"
/dev_stage="Adult"
/note="Organ: breast; Vector: puc18; Site_1: SmaI; Site_2:
SmaI; A mini-library was made by cloning products derived
from ORESTES PCR (U.S. Letters Patent application No. 196
,716 - Ludwig Institute for Cancer Research) profiles
into the puc 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions."
BASE COUNT 198 a 89 c 89 g 174 t
ORIGIN
Query Match 36.5%; Score 499.6; DB 9; Length 550;
Best Local Similarity 97.6%; Pred. No. 4.5e-74;
Matches 539; Conservative 0; Mismatches 9; Indels 4; Gaps 3:
Qy 171 gtggatatttcacaaaacacattcaatgttttaaacactattttttgaagaacaaaa 230
Dy 3 CGGGGATATGTTTACAAAACACATTCATGTTTAAACACTATTTTTCGAACAACAAA 62
Qy 231 tatatttaaaattgttgccttcttaaaagccatttccctcccaagctcaacttttaa 290
Dy 63 TATA--TGAAAATGTTTCCTTCTTAAAGGCCATTTCCTCCCAAGTCTAANCITTTGTA 120
Qy 291 ttgatattaagaatgaagttattttgtacaattctagtttaaacagcagaatagcactag 350
Dy 121 TTGATATTAGCAATTCAGATTATTTTGTACAACTAGTTTAAACAAGCAGATTAGCAGTAG 180
Qy 351 gcagaataaaaaattgcacagcagctatgcataattttccaagaatacattcttaaatcag 410
Dy 181 GCAGAAATAAAAAATTGCACAGACCGATGCAATTTTCCAAAGATGATCTTTTAATTCAG 240
Qy 411 ttttcagcttccaaagattggtgcccataatagacttaaacatataatgatgagctaaaa 470
Dy 241 TTTTACCTTCGAAGATGGGTGGCCCATATAGACTTTAAACATATATATGATGGCT-AAA 299
Qy 471 aaataagtatcagaaattgtataaaagggaatgttaagtccacactcaat:tcataaaa 530

```

---

```

Db 300 AAAATAAGTATACGAAAATGTAAAAAGGAANTGTAAGTCCACTCTCAATCTCATNAAG 359
Qy 531 gtgaagttaagctaaacacaaaataaaatgtaggtctctttttttctgttccgtttat 590
Dy 350 GTGAGAGTAAGGATGCTTAACGCAAAATAAATGTAAGTCTCTTTTCTATTTCGGTTAT 419
Qy 551 catgaattctgtcttctttgatctatcccttagggtttacccatttaagttagaggtgtaag 650
Dy 420 CATGCAATCTGCTTCTTTGATATGCTTAGGTTTACCAITTAAGTTAGAGGTGTAAAG 479
Qy 651 caatgtggaataaaaaattgaataataacacattctgtcatcttcttcaattcacaattgcgg 710
Dy 480 CAATGTGGGAATGAATAATCATCAATATACACCTTGTCATTTCAITTCAAATCGC-GG 538
Qy 711 ctggaataacttc 722
Dy 539 CTGGAACACTTC 550

RESULT 4
BE872674
LOCUS 501451222F1 NIH_MGC_65 Homo sapiens cDNA clone IMAGE:3855125 5
DEFINITION mRNA sequence.
ACCESSION BE872674
VERSION BE872674.1 GI:10321450
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 905)
NIH-MGC http://img.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLMS581 row: 1 column: 06
High quality sequence stop: 582.
Features
Location/Qualifiers
1..905
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3855125"
/clone_lib="NIH_MGC_65"
/tissue_type="adenocarcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: colon; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.8 kb. Library constructed by Life
Technologies."
BASE COUNT 277 a 245 c 143 g 240 t
ORIGIN
Query Match 28.8%; Score 393.4; DB 10; Length 905;
Best Local Similarity 83.0%; Pred. No. 1.9e-56;
Matches 545; Conservative 2; Mismatches 93; Indels 17; Gaps 8:
Qy 590 tcatgcaattctcttctttgatctatcccttagggtttacccatttaagttagaggttgaat 649
Dy 762 ICAAGCAATGTCGCTCTTTAGATCTCTTAGGGTTACCCCTTTTAA--TTAAGTTGTAAAG 705
Qy 650 gcaatggtgggaatgaataattgaataataacacacttg-----tcatttcatttcaaat 704

```

```
Db 704 ACAATTGTTGGGATTTAAATGGTCGAGATGTGCGCCCTGTCCTGTCAGTTTCTATT 645
Oy 705 tccgggtcagaactcccaaaaaagggtta--ggcatgaagaaaaaataatcmaatca 761
Db 644 TCGGGCTGGGAAGCGTCGCCGATAAAGGGGTAGGCTTGTAGAGTGTAGTAAACGAATCA 585
Oy 762 gaacctcttcaggggttgkktctgatatggcagacargatacaagaatcccccacagaga 821
Db 584 GAACCTCTTCAGGGGTTT--GTTCGTGATATGGCAGACAAAGTACAAAGTCCACCAGGAGA 527
Oy 822 tggagcaattcaaaataagggtaattgggtcagcaaggattatttgcacatgagacaga 881
Db 526 TCGAGCAATTCAGATTAAGGGTAATGGGCTGACTAGGTATTATGSCCAGCATGGGACAGA 467
Oy 882 atgagcaacagcgtcaaaagtttttgattatatagcaccctagagctctctgatagaga 941
Db 466 ATGAGCAGAGGCTGAAAAGTGTGTGGATTATATAGCACCTAGAGTCTCTCAITGAGGGA 407
Oy 942 atttttgtagtcaaacatcacctcaacttccaagggaataatcttccaggtagcctaadc 1001
Db 406 ATTITITGTTAGTCAACATACGCTAAACITCCAAAGGGGAAATCTTTCAGGTAGCCTAAGC 347
Oy 1002 ttgctttctadagtgatgagtgctgctgctgctgctgctgctgctgctgctgctgct 1061
Db 346 TTGCTTTCTAGAGTCAATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 288
Oy 1062 tgtacaagtgagaaagactagagagagaagatttttaagctctttagcagagcattttat 1121
Db 287 TGTAC--AGTGAGAAAGACTAGAGAGAAAGATTITAGTCTGTTTAGCAGAGGCAATTTTAT 229
Oy 1122 ctgctgacatgagatcaatcttctgctcctctacccctacccaggaaggcgaatcccaa 1181
Db 228 CTGGTGACATCGATCAATATTTCTGATCCCTATATCCCGAGGAAGGCAAA--TCCCA 171
Oy 1182 agaatgtgttagcaaaatggctgctgctgctgctgctgctgctgctgctgctgctgct 1238
Db 170 AGAAATGTGTAGC--AAATGGTGATGCTATCAATATGCTATGCTATGCTATGCTATG 115

RESULT 5
A1085849
LOCUS ox47f12.s1 Soares_total_fetus_Nb2Hf8_9w Homo sapiens cDNA clone
DEFINITION IMAGE:1659503 3', mRNA sequence.
ACCESSION A1085849
VERSION A1085849
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 409)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP).
JOURNAL Tumor Gene Index
COMMENT Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapsb-r@mail.nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1869 Std Error: 0.00
Seq Primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 398.
Location/Qualifiers
1..409
/db_xref="taxon:9606"
/organism="Homo sapiens"
/clone="IMAGE:1659503"
/clone_lib="Soares.Total_fetus_Nb2Hf8_9w"
/dev_tag="8-9 weeks"
/lab_host="DH10B"
/note="vector: p7T3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not 1; Site_2: Eco RI; 1st strand cDNA
```

was prepared from mRNA obtained from pooled 8-9 week (total) fetus material with a Not I - oligo(dI) primer [5'-TGTTACCAATCGAATGGAGCGGCCCTTAATTTTCTTTTCTTTT 3']. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo. \*

BASE COUNT 158 a 58 c 59 g 134 t  
ORIGIN

Query Match 28.4%; Score 388.8; DB 9; Length 409;

Best Local Similarity 99.5%; Pred. No. 1.6e-55;

Matches 390; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 82 tttttattgacagcaattcaggctttattattgagcaatgaaaaacagctaaaaacttaatt 141

Db 1 TTTTITATGATCAGCAATTCAGGCTTTATTATTGAGCAATGAAAACAGCTTAAACCTTAAT 60

Oy 142 tccaagcatgtgattgaaagtttgcaagtggaatattttcacaaaaacacattcaatg 201

Db 51 TCCAAGCATGTGAGTTAAAGTTTGCAAGTGGGATATTGTCACAAAACACATTCGAATG 120

Oy 202 tttaacacactattttgaagaacaaaatatatttaaaattgttgccttcttaaaagcc 261

Db 121 TTTAAACACTATTATTGGAAGAACAAATATATTAAAAATTGTTTCCTTCTAAAAAGCC 180

Oy 262 catttcctccaaagctaaactttgtaattgtatattgaagcaatgaagttattttgtaca 321

Db 181 CATTTCCCTCCNAGCTTAAACTTTTGTAATTGATATTAGCAATGAAGTTATTGTGACA 240

Oy 322 atctagttaaacacagcaatagcactaggcagaataaaaaattgcacagcgtatgcaa 381

Db 241 ATCTAGTTAAACAAGCAGCAATAGCACTAGGCAGCAATAAAAAATTCACAGACGTATGCCAA 300

Oy 382 ttctccagatagcatctttaaattcagtttccagcttccaaagattggtgcccataa 441

Db 301 TTTTCCAAGATAGCATCTTTTAATTTCAGTATTCAGCTTCCAAAGATTGTTGGGCATAA 360

Oy 442 tagacttaacataatgatggctaaaaaaa 473

Db 361 TAGACTTAAACATATAATGATGGCTAAAAAAA 392

RESULT 6

A1867547

LOCUS wb78c04.x1 NC1-CGAP\_Pr28 Homo sapiens cDNA clone

DEFINITION mRNA sequence.

ACCESSION A1867547

VERSION A1867547.1

KEYWORDS GI:5540656

SOURCE EST.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 364)

AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP).

JOURNAL Tumor Gene Index

COMMENT Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Email: cgapsb-r@mail.nih.gov

Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.

Emmert-Buck, M.D., Ph.D.

CDNA Library Preparation: M. Bento Soares, Ph.D.

DNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 903 Std Error: 0.00



```

Seq primer: -40up from Gibco
High quality sequence stop: 362.
Location/Qualifiers
1. 354
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2311782"
/clone_lib="NCI_CGAP_Pr28"
/sex="male"
/dev_stage="adult"
/lab_host="DH10B"
/note="Organ: Prostate; Vector: p733D-Pac (Pharmacia)
with a modified polylinker; Plasmid DNA from the
normalized library NCI_CGAP_Pr22 was prepared, and ss
circles were made in vitro. Following HAP purification,
this DNA was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from a pool
of 5,000 clones made from the same library (clonelids
985608-986759, 1101192-1101959, and 1217928-1220615).
Subtraction by Bento Soares and M. Fatima Bonaldo."
BASE COUNT 134 a 55 c 52 g 122 t
ORIGIN
Query Match 25.7%; Score 352; DB 9; Length 364;
Best Local Similarity 99.7%; Pred. No. 2.4e-49;
Matches 363; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
Oy 85 ttattgacgaattcaggctttattattgagcaatgaaacagctaaacattattcc 144
Db 1 tttattgacgaattcaggctttattattgagcaatgaaacagctaaacattattcc 50
Oy 145 aagcatgttagttaaagttgcaagtggtgatttgcacaaacacattcaatgttt 204
Db 61 AAGCATGTGTAGTTAAAGTTTGCAAGTGGGATATTGTTCAACAAACACATTCAATGTT 120
Oy 205 aacactatttattgaagaacaaatatatttaaattgttgccttcaaaaagcccat 264
Db 121 AAMACATATTATTGAAGAACAAATATATTAAATTTGTTGTTCTTAAAGCCCAT 180
Oy 265 tccccccaagctcaactttgtaatttgatattagcaatgaagtattttgtacaatc 324
Db 181 TTCCCTCCAGTCTAAACTTTGTAATTTGATATTAGCAATGAAGTATTTTGTACAAATC 240
Oy 325 tagttaaacaagcagatagcactaagcagaataaaattcacagacgtatgaacttt 384
Db 241 TAGTTAAACAAGCAGAAATAGCATTAGCAGATAAAATTTGCACAGACGATGCANATTT 300
Oy 385 tccaagatagcattctttaaattcagtttgcagcttcccaaga-ttgggtgcccataata 443
Db 301 TCCAGATACCATTCITTTAAATTCAGTTTTCAGCTTCCAAAGATTGTTGTCGCAATAA 360
Oy 444 gact 447
Db 361 GACT 364

RESULT 7
AG030921 AG030921 746 bp DNA linear GSS 01-NOV-2001
LOCUS Pan troglodytes DNA, clone: PTB-003H07.F, genomic survey sequence.
DEFINITION AG030921
ACCESSION AG030921
VERSION AG030921.1 GI:15557794
KEYWORDS GSS (genome survey sequence).
SOURCE Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male
BAC library clone:PTB-003H07.F.
ORGANISM Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Pan.
1 (sites)
Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Totoki,Y., Watanabe,H. and Sakaki,Y.
BAC end sequences of library PTB

REFERENCE
AUTHORS
TITLE

JOURNAL Unpublished
REFERENCE 2 (bases 1 to 746)
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Totoki,Y., Watanabe,H. and Sakaki,Y.
DIRECT SUBMISSION
JOURNAL Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-chou,Tsukumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:chimbesc@sc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
Clones are derived from the chimpanzee BAC library PTB This BAC end
was generated during the R&D process and may have higher chance of
clone tracking errors.
PRIMERS
Sequencing: -21M13
LIBRARY Vector : pKS145
R.Site 1 : SacI
R.Site 2 : SacI.
Location/Qualifiers
1. 746
/organism="Pan troglodytes"
/db_xref="taxon:9598"
/clone="PTB-003H07.F"
/sex="male"
/cell_type="Lymphoblast"
/clone_lib="PTB Chimpanzee Male BAC Library"
BASE COUNT 213 a 173 c 135 g 224 t
ORIGIN
Query Match 17.1%; Score 234.4; DB 12; Length 746;
Best Local Similarity 97.5%; Pred. No. 7.4e-30;
Matches 238; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
Oy 82 ttttttattgacgaattcaggctttattattgagcaatgaaacagctaaacattta 141
Db 503 TTTTCTATTGATGAGATTTCAGGCTTTATTATTGAGCAATGAAACAGCTTAAACITTA 562
Oy 142 tccaagcatgttagttaaagttgcaagtggtgatttgcacaaacacattcaatg 201
Db 563 TCCAGCATGTGTAGTTAAAGTTTGCAAGTGGGATATTGTTCAACAAACACATTCAATG 622
Oy 202 ttaaacactatttatttgagaacaaatatatttaaattgttgccttcaaaaagcc 261
Db 623 TTAAACACTGTGTTATTATTGAAGAACAAATATATTAAAAATTTGTTCTTAAAGCC 682
Oy 262 catctccctcgaagtctaaacttttgatatttgatattgaagcaatgaagtattttgtaca 321
Db 583 CATTTCCCTCCAGCTTAAACTTTGTAATTTGATATTGATTAAGCAATGAAGCTATTATTACA 742
Oy 322 atct 325
Db 743 GTCT 745

RESULT 8
BM391501 BM391501 661 bp mRNA linear EST 17-JAN-2002
LOCUS UI-R-DY0-ckr-e-06-0-UI.sl UI-R-DY0 Rattus norvegicus cDNA clone
DEFINITION UI-R-DY0-ckr-e-06-0-UI 3', mRNA sequence.
ACCESSION BM391501
VERSION BM391501
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 661)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery

```







Soares Lab Clone Distribution: clones will be available through Research Genetics (www.resgen.com) The following repetitive elements were found in this cDNA sequence: 413-191, >R1\_MM#SINE/Alu 436-491, >PB1D10#SINE/Alu 511-570, >ID3#SINE/ID  
Seq primer: M13 Forward  
POLYA#res.

FEA#URES  
source  
Location/Qualifiers  
1. 733  
/organism="Rattus norvegicus"  
/strain="Sprague-Dawley"  
/db\_xref="taxon:10116"  
/clone="UI-R-DY0-ckg-k-16-0-UI"  
/clone\_lib="UI-R-DY0"  
/dev\_stage="adult"  
/lab\_host="DH10B (Life Technologies)"  
/note="vector: p713D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; UI-R-DY0 is a non-normalized Rat cartilage library (RC) constructed in p1317 PAC vector according to the procedure described by Bonaldo, Lennon & Soares (Genome Research genome 6: 791-806, 1996). The oligonucleotide used to prime first strand synthesis contained the sequence tag CTAATGACG between the Not I cloning site and dH18 stretch. The Rat cartilage tissue was provided by Dr Jeff Stevens at the University of Iowa.  
TAG\_LIB=UI-R-DY0  
TAG\_TISSUE=cartilage  
TAG\_SEQ=CTAATGACG  
BASE COUNT 214 a 149 c 137 g 232 t 1 others  
ORIGIN

Query Match 7.3% Score 99.4; DB 10; Length 733;  
Best Local Similarity 60.1%; Pred. No. 2.6e-07;  
Matches 224; Conservative 0; Mismatches 136; Indels 13; Gaps 3;  
Qy 211 tatatttgaagacacaaatattttaaattgttgcctctataaaagccattccct 270  
Db 10 TTTTATTTGAAGACAGAAATATATTCACACITGTTTATCTCTACAAACCCATTT CCT 68  
Qy 271 caagctcaaaatttgtaattgataattaagcaatgaattta-ttttgacatcaatt 329  
Db 69 CAAGCCTTAATTTATGCTGCTGTGTTAAGCAATGAAGTTATTTTATTAANACCTAGTT 128  
Qy 330 aaacagcagatagcactaggcagagaataaaaaattgcacagacgtatgcgaattttccaa 389  
Db 129 AAGCAATTTAATGGGAATAGCATCGTTAGCAATAAAAAATTGCACACATCGGTTACTCTCA 188  
Qy 390 gatagcattctttaaattcagtttccagtttccaaagattgattcccaataagactta 449  
Db 189 AATGATGATTTATTTAGTGTCAACTTTCAGCTTCTAGTGACAAATTCCTTCAIAGACCTGA 248  
Qy 450 aacataaatgatggctaaataaaataaagtatacagaaaatgtaaaaaaggaatgtaagt 509  
Db 249 AATGCATAATGATAGTTGAAGAAATATATATGTAAAGAAGTAA-----AATA 297  
Qy 510 ccaactcgaatcctataaaggtagagtagaagtagctgaagcaaaaataaatgtagggttc 569  
Db 298 CATTCGTGTCATATATAAAGGAGACCGTAGGATACACACGCTGAAGGAAPCGGCATTT 357  
Qy 570 ttttttctctgtt 582  
Db 358 TCTTTTGTGCTT 370

RESULT 13  
LOCUS AU177437/c 535 bp mRNA linear EST 21-MAR-2001  
DEFINITION AU177437 OLHNI cell line cDNA library (OLB) Oryzias latipes cDNA clone OLB01.01h similar to pirl158524; ribosomal protein L34 human, mRNA sequence.  
ACCESSION AU177437  
VERSION AU177437.1 GI:13426273

KEYWORDS  
SOURCE  
ORGANISM

EST.  
Japanese medaka.  
Oryzias latipes  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha; Belontiiformes; Adrianichthyidae; Oryziinae; Oryzias.  
REFERENCE 1 (bases 1 to 535)  
AUTHORS Naruse.K., Mitani.H. and Tanaka.M.  
TITLE Medaka EST Project in University of Tokyo (2001)  
JOURNAL Unpublished (2001)  
COMMENT Contact: Kiyoshi Naruse  
Department of Biological Sciences  
Graduate School of Science, University of Tokyo  
Hongo 7-3-1, Bunkyo-ku, Tokyo 113-0033, Japan  
Tel: 81-3-5841-4443  
Fax: 81-3-5841-4410  
Email: naruse@biol.s.u-tokyo.ac.jp  
This clone was isolated from OLHNI cell line cDNA library (OLB) 5' end sequences.

FEATURES  
source

Location/Qualifiers  
1. 535  
/organism="Oryzias latipes"  
/strain="HNI"  
/db\_xref="taxon:8090"  
/clone="OLB01.01h"  
/clone\_lib="OLHNI cell line cDNA library (OLB)"  
BASE COUNT 150 a 135 c 147 g 103 t  
ORIGIN

Query Match 6.8% Score 92.4; DB 9; Length 535;  
Best Local Similarity 83.3%; Pred. No. 4.4e-06;  
Matches 105; Conservative 0; Mismatches 21; Indels 0; Gaps 0;  
Qy 1 ccaatgaacgcgcgtaatacagactcactataggcggaattgggtaccggcccccctcg 60  
Db 499 CCAGTGACGCGCGTATACGACTCACATAGGCGGAATGGGTACCGGCCCCCTCG 440  
Qy 51 agcgccgcctcttt 120  
Db 439 AGCGCGCGCGGACCGTTTTTTTTTTTTTTTTCATTCAGGTGTGTTTATTATTGGCA 380  
Qy 121 tgaaaa 126  
Db 379 GTACAA 374

RESULT 14  
LOCUS AG054901

DEFINITION Pan troglodytes DNA, clone: PTB-040L14.F, genomic survey sequence.  
ACCESSION AG054901  
VERSION AG054901.1 GI:16592344  
KEYWORDS GSS: GSS (genome survey sequence).  
SOURCE Pan troglodytes male lymphoblast DNA, clone\_lib:PTB-Chimpanzee Male  
ORGANISM BAC Library clone:PTB-040L14.F.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.  
REFERENCE 1 (sites)  
AUTHORS Fujiyama.A., Hattori.M., Toyoda.A., Taylor.T.D., Yada.T., Totoki.Y., Watanabe.H. and Sakaki.Y.  
TITLE BAC end sequences of Library PTB  
JOURNAL Unpublished  
AUTHORS 2 (bases 1 to 658)  
TITLE Fujiyama.A., Hattori.M., Toyoda.A., Taylor.T.D., Yada.T., Totoki.Y., Watanabe.H. and Sakaki.Y.  
JOURNAL Direct Submission  
AUTHORS Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:chimpes@gsc.riken.go.jp, URL:http://hqp.gsc.riken.go.jp/,

COMMENT Tel:81-45-503-9111, Fax:81-45-503-9170)  
Clones are derived from the chimpanzee BAC library PTB This BAC end  
was generated during the R&D process and may have higher chance of  
clone tracking errors.

PRIMERS  
Sequencing: -21M13

LIBRARY  
Vector : pKS145  
R.Site 1 : SacI  
R.Site 2 : SacI  
Location/Qualifiers  
1. .658  
/organism="Pan troglodytes"  
/db\_xref="taxon:9598"  
/clone="PTB-040L14.F"  
/sex="male"  
/cell\_type="lymphoblast"  
/clone\_lib="PTB Chimpanzee Male BAC Library"  
BASE COUNT 184 a 157 c 127 g 189 t 1 others

Query Match 6.7%; Score 92; DB 12; Length 658;  
Best Local Similarity 83.7%; Pred. No. 4.7e-06;  
Matches 144; Conservative 0; Mismatches 15; Indels 13; Gaps 3;  
Qy 82 ttttttattgacgaattcagcgtttatttattgagcaatgaaacacagctaaacttaatt 141  
Db 487 TTTTCTATTGATCAGATTCAGGCTTTATTATTGAGCAATGAAAACAGCTAAACTTAAI 546  
Qy 142 tccaagcatgtgtagttaaagtttgc-----aaagtggtattgtttcacaaaaaacac 193  
Db 547 TCCAGCATGTGTAGTTAAAGTATGCGACGACGGGGAGATATTGTTACAAAACCA 505  
Qy 194 at-tcaatgtttaaacac-----tattttttgaagacaataatatattaaa 240  
Db 607 AINTCAATGGTTAAACACATGTTGTTATTATGAGACAAAATATATTAAA 658

RESULT 15  
AW790437  
LOCUS  
DEFINITION AW790437 677 bp mRNA linear EST 01-MAY-2001  
C01755-F Lambda zap, Stratagene Blumeria graminis f. sp. hordei  
CDNA clone C01755 similar to catalase, mRNA sequence.  
ACCESSION  
VERSION AW790437.1 GI:13902034  
KEYWORDS  
SOURCE  
ORGANISM  
Blumeria graminis f. sp. hordei.  
Blumeria graminis f. sp. hordei  
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Leotiomycetes;  
Erysiphales; Erysiphaceae; Blumeria.  
REFERENCE  
AUTHORS 1 (bases 1 to 677)  
Thomas,S.W., Rasmussen,S.W., Glaring,M.A., Rouser,J.A. and Oliver  
,R.P.

TITLE Gene identification in the fungal pathogen Blumeria graminis by  
expressed sequence tag analysis  
JOURNAL unpublished (2000)  
COMMENT Contact: Rasmussen,S.W.  
Department of Yeast Genetics  
Carlsberg Laboratory  
10 GL Carlsbergvej, DK-2500, Copenhagen, Denmark  
Tel: 45 3327 5230  
Fax: 45 3327 4756  
Email: swr@rc.dk  
High quality sequence stop: 677  
POLY(A)-No. Location/Qualifiers

FEATURES  
source  
1. .677  
/organism="Blumeria graminis f. sp. hordei"  
/db\_xref="taxon:62688"  
/clone="C01755"  
/clone\_lib="Lambda zap, Stratagene"  
/cell\_type="conidia"

BASE COUNT 162 a 179 c 191 g 143 t 2 others  
ORIGIN  
/lab\_host="Hordeum vulgare"  
Query Match 6.5%; Score 88.8; DB 9; Length 677;  
Best Local Similarity 93.0%; Pred. No. 1.6e-05;  
Matches 93; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
Qy 1267 gactagtggatccactactagttcttagagcgcgcgcccaccgcggtggagctccagctttgt 1326  
Db 19 GCCCGGGGATCCACTAGTTCTAGAGCGCGCCGCCACCGCGTGGAGCTCCAGCTTTTGT 78  
Qy 1327 tcccttttagtgagggttaattgacgcgcttgcgtaatacat 1366  
Db 79 TCCCTTTTAGTGAGGGTTAAATTCGAGCTTGGCGTAATCAT 118  
Search completed: April 29, 2002, 17:16:55  
Job time: 13214 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 29, 2002, 17:16:55 ; Search time 6783.7 Seconds  
(without alignments)  
1838.405 Million cell updates/sec

Title: US-09-248-178-52  
Perfect score: 924  
Sequence: 1 caaaggnacaggaacagctt.....aagggcaaaaaaaaaaaaaa 924

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estnu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_estl:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_gss:\*  
13: em\_gss\_hum:\*  
14: em\_gss\_inv:\*  
15: em\_gss\_pin:\*  
16: em\_gss\_vrt:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	505.6	54.7	519	9	A1917901
C 2	453	49.0	468	9	AW469374
C 3	449.2	48.6	461	9	A1801231
C 4	340.4	36.8	450	10	R78938
C 5	323	35.0	505	10	BG007631
C 6	258	27.9	414	9	AW070731
C 7	98.2	10.6	215	9	AA228748
C 8	61.2	6.6	321	10	BG009951
C 9	59.6	6.5	579	12	AG057824
C 10	58.4	6.3	393	9	A1937404
C 11	57.8	6.3	433	12	AQ115887
C 12	57.6	6.2	508	12	AQ598884
C 13	57.2	6.2	420	12	AQ153688
C 14	56.8	6.1	304	9	AA621843
C 15	56.8	6.1	372	9	A1056137
C 16	56.8	6.1	518	12	AQ885032
C 17	56.6	6.1	462	12	AQ226207

18	56.4	6.1	459	12	AQ210411
C 19	56.2	6.1	481	12	AQ001555
C 20	55.8	6.0	597	12	AQ554308
C 21	55	6.0	658	12	AG106298
C 22	54.8	5.9	360	9	A1081708
C 23	54.8	5.9	400	12	AQ789932
C 24	54.8	5.9	410	10	N88393
C 25	54.6	5.9	355	12	AQ051860
C 26	54.6	5.9	467	12	AQ580989
C 27	54.6	5.9	524	12	AQ695826
C 28	54.4	5.9	453	12	AQ129634
C 29	54.2	5.9	624	9	AW004000
C 30	54.2	5.9	842	12	AQ781743
C 31	53.8	5.8	754	12	AQ636028
C 32	53.8	5.8	869	10	BM007637
C 33	53.6	5.8	390	9	A1269041
C 34	53.6	5.8	392	9	AA463632
C 35	53.6	5.8	472	9	A1458879
C 36	53.6	5.8	480	9	A1580056
C 37	53.6	5.8	522	10	BE677940
C 38	53.6	5.8	561	10	BF196874
C 39	53.6	5.8	698	12	AG093829
C 40	53.2	5.8	440	12	AQ167953
C 41	53.2	5.8	999	9	AL547605
C 42	52.8	5.7	471	12	AQ123040
C 43	52.8	5.7	651	12	AG035074
C 44	52.4	5.7	503	12	AQ568884
C 45	52.4	5.7	538	9	AL042165

## ALIGNMENTS

RESULT 1  
A1917901/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

A1917901 519 bp mRNA linear EST 28-JUL-1999  
t214d04.x1 NCI\_CGAP\_Ut2 Homo sapiens cDNA clone IMAGE:228551 3',  
mRNA sequence.  
A1917901 GI:5637756  
EST.  
human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 519)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP).  
Tumor Gene Index  
Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Email: [cgapsb@mail.nih.gov](mailto:cgapsb@mail.nih.gov)  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
Emmert-Buck, M.D., Ph.D.  
CNA Library Preparation: Life Technologies, Inc.  
CNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
[www-bio.llnl.gov/bbrp/image/image.html](http://www-bio.llnl.gov/bbrp/image/image.html)  
Seq primer: -40UP from Gibco  
High quality sequence stop: 397.  
Location/Qualifiers  
1. 519  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:228551"  
/tissue\_type="moderately-differentiated endometrial  
adenocarcinoma, 3 pooled tumors"  
/lab\_host="PH10B"  
/note="organ: uterus; vector: pCMV-SPORT6; Site:1: SalI;  
Site:2: NotI; Cloned unidirectionally. Primer: Oligo dr.

18 56.4 6.1 459 12 AQ210411  
C 19 56.2 6.1 481 12 AQ001555  
C 20 55.8 6.0 597 12 AQ554308  
C 21 55 6.0 658 12 AG106298  
C 22 54.8 5.9 360 9 A1081708  
C 23 54.8 5.9 400 12 AQ789932  
C 24 54.8 5.9 410 10 N88393  
C 25 54.6 5.9 355 12 AQ051860  
C 26 54.6 5.9 467 12 AQ580989  
C 27 54.6 5.9 524 12 AQ695826  
C 28 54.4 5.9 453 12 AQ129634  
C 29 54.2 5.9 624 9 AW004000  
C 30 54.2 5.9 842 12 AQ781743  
C 31 53.8 5.8 754 12 AQ636028  
C 32 53.8 5.8 869 10 BM007637  
C 33 53.6 5.8 390 9 A1269041  
C 34 53.6 5.8 392 9 AA463632  
C 35 53.6 5.8 472 9 A1458879  
C 36 53.6 5.8 480 9 A1580056  
C 37 53.6 5.8 522 10 BE677940  
C 38 53.6 5.8 561 10 BF196874  
C 39 53.6 5.8 698 12 AG093829  
C 40 53.2 5.8 440 12 AQ167953  
C 41 53.2 5.8 999 9 AL547605  
C 42 52.8 5.7 471 12 AQ123040  
C 43 52.8 5.7 651 12 AG035074  
C 44 52.4 5.7 503 12 AQ568884  
C 45 52.4 5.7 538 9 AL042165



Average insert size 1.85 kb. Life Technologies catalog # 11539-012*									
BASE COUNT	154 a	104 c	107 g	154 t					
ORIGIN									
Query Match	54.7%: Score 505.6; DB 9; Length 519;								
Best Local Similarity	99.2%: Pred. No. 1.le-103;								
Matches 516; Conservative	2: Mismatches		1: Indels		1: Gaps				
Oy	391 aaatatagctactctattgtcattattattacttctgtttgaagctaaagacctcacata 450								
Db	519 AAATATAGTACTTACTATTGTCAITATTATTACTTGTITGAAGCTTAAAGAGCTCACATA 460								
Oy	451 gaattccattccagccaccagacaagagctctgaagtttcttagtttgaagagctattaaa 510								
Db	459 GAATCCATCCAGCCACCAGACAGAGCTCTGAGTTTTCTTAGTTTGGAGAGCTATTAAA 400								
Oy	511 taacaactctagctgaattctctatacttctgtttatgttcaagttaactgggctcagcat 570								
Db	399 TAACAAGCTCTAGTGTCAATTGTATCTTGTATGTGTCAAAGT-AC'TGGGCTCAGCATTTT 341								
Oy	571 acattcattgtctcttaagttctgaagctgaagcaggaagctatgattattatgacta 530								
Db	340 ACATTCATTGCTCTTTTAGTTCTAGCAATGTGACGAGGAGCAATGATTATTATGACTA 281								
Oy	631 cataaatgaagaattgaggtccagatatacttaagtaaatctctccagggtcacacagcta 590								
Db	280 CATAAATCAAGAAATTTAGGCTCAGATACATTAAAGTAATTTCTCCAGGGTCCACAGCTA 221								
Oy	691 gaactggcaagcctggagttgaatccatgatcttccagcattggaagaatcataaagttaa 750								
Db	220 GAACTGCGAAAGCTGGGATTGATCCACGATCTTCAGCATTTGANGAATCATATAAIGTAA 161								
Oy	751 ataactgaaggcctttctcagaagagctcctggtgtctgcacaaacccactgacct 810								
Db	160 ATAACTGAAGGCTTTTCTCAGAAGAGCTCTGTGTGTTGCAGCAACCCACTAGCAGC 101								
Oy	811 tgtctctcagggggaacatctgtggcctgggaatcactgcacgtcgcaagagatgttg 870								
Db	100 TGTCTCTACAGGGGAACATCTGTGGCCTGGGAATCACTGCACGCTCGCAGAGATGTTG 41								
Oy	871 ctctctgataattatttctctcctcagttggtgtaaggca 910								
Db	40 CTCTGTGATGAATTAATGTTCTGTCTGAGTGTGTGAAGGCA 1								
RESULT 2									
AW459374/c									
LOCUS	AW459374		458 bp		mrna		linear		EST 24-FEB-2001
DEFINITION	hc82h01.x1 NCI_CGAP_Gas4 Homo sapiens cDNA clone IMAGE:289249 3'								
ACCESSION	mrna sequence.								
VERSION	AW459374								
KEYWORDS	AW459374.1 GI:7039480								
SOURCE	EST.								
ORGANISM	human.								
REFERENCE	Homo sapiens								
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;								
TITLE	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.								
JOURNAL	1 (bases 1 to 458)								
COMMENT	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap. National Cancer Institute, Cancer Genome Anatomy Project (CGAP). Tumor Gene Index Unpublished (1997) Contact: Robert Strausberg, Ph.D. Email: cgapb-r@mail.nih.gov Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: Life Technologies, Inc. cDNA Sequencing by: Greg Lennon, Ph.D. DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: NCI-CGAP clone distribution information can found through the I.M.A.G.E. Consortium/LLNL at:								

RESULT	4
R78938/c	
LOCUS	R78938
DEFINITION	y187e07.s1 Soares placenta Nb2HP Homo sapiens cDNA clone IMAGE:146244 3', mRNA sequence.
	450 bp
	mRNA
	linear
	EST 09-JUN-1995

TATTTGACTACATAATGAAGAAATTCAGGCTCAGATACATTAAAGTAATTTCTCCAGGGT 209  
 TACACAGCTAGAAGCTGGCAAG·CCTGGGATGATCCATGATCTCCAGCATTTAAAGAAAT 739

TATTTGACTACATAATGAAGAAATTCAGGCTCAGATACATTAAAGTAATTTCTCCAGGGT 209  
 TACACAGCTAGAAGCTGGCAAG·CCTGGGATGATCCATGATCTCCAGCATTTAAAGAAAT 739

```

Db 208 CACACAGCTAGAAACTGCAGAGCCCTGGGATGATCCCATGATCTTCCACCATTTGAAGAA 149
Qy 740 cataaatgaaataaactgcaagcccttttctcagaagagct-cctgggtgctgagcaacaa 798
|||||
Db 148 CATAAATGAAATAACTGCAAGGCCCTTTCTCTCAGAAGAGCTCCCTGGTCTTGACCAAA 89
Qy 799 ccactagacactgttctctacaggggaacatctgtggccctgggaatacactgcacgtcg 858
|||||
Db 88 CCACATAGACACTGTCTCTACAGGGGAACATCTGTGGGCCCTGGGAATCACTGCAAGTCG 29
Qy 859 caagatgtgtctctgtatgaattatt 885
Db 28 CAAGAGATGTGTCTCTGATGAATTATT 1

RESULT 5
LOCUS BGO 7631
DEFINITION QV4-CNO250-281100-608-f05 GN0250 Homo sapiens cDNA, mRNA sequence.
ACCESSION BC007631
VERSION BC007631.1 GI:12452005
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 505)
AUTHORS Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carrvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL PROC Natl. Acad. Sci. U.S.A. 97 (7), 3491-3495 (2000)
MOLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl-QV4&t2-QV4-GN0250-
281100-608-f05&t3-2000-11-28&t4-1)
Seq primer: puc 18 forward
High quality sequence start: 17
High quality sequence stop: 505.
FEATURES
Location/Qualifiers
1..505
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="GN0250"
/der_stage="Adult"
/note="Organ: Placenta normal; Vector: puc18; Site_1: SmaI
; Site_2: SmaI; A mini-library was made by cloning
products derived from ORFESTES PCR (U.S. Letters Patent
application No. 196,716 - Ludwig Institute for Cancer
Research) profiles into the pUC 18 vector. Reverse
transcription of tissue mRNA and cDNA amplification were
performed under low stringency conditions."
BASE COUNT 166 a 101 c 96 g 142 t
ORIGIN

```

```

Qy 1 caaaggnacagaagaacagctcttgaagaagtactgcatcctcctcagcaggaacaccccttt 60
|||||
Db 49 CCAAAGGACAGGACAGAGCTTG-AAAGTACTTTCAT-CTTCTCTGAGGGACACGCCCTTT 105
|||||
Qy 61 gcctccaaagcaatagaaattttaaagatttncactcagagaagggggnccacgtttnart 120
|||||
Db 107 GCCTCAAAAGCAATAGCAAAATTTAAAGATTT-CACTGAGAAGGGGCC--ACGTTTAAAG 163
|||||
Qy 121 tntnaatgntcargnanarntccttncacaaatgncrntcactnactnactnactnactnactn 180
|||||
Db 154 TTATAATGTC TAAGGAAGACACTTCAAAATGCGACATACATACATACATAGAAATTTGGG 223
|||||
Qy 181 tncgqrntcmgnactatntcaggtttgaaaaactggatctgcccacttatcagttatgtg 240
|||||
Db 224 TTCCAGAGTTCAGA--CATATCTAGCTTTTGAAGCTGGATCTGCCACTTACCAGTTATGTG 281
|||||
Qy 241 accttaagaagaactcgttaatttctcagagcctcagttctcctcctataagttggaggt 300
|||||
Db 282 ACCITTAAGAACTCCGTTAAATTTCTCAGAGCCTCAGTTTCTTGTCTATAAGTTGGAGT 341
|||||
Qy 301 aaatataactatcatcttttccaaaggttgatggaacattaatgagtgaaatgacag 360
|||||
Db 342 AATATTAACTATCATTTTTTCCAAAGGATTGATGTGAACATTAATGAGTGAATGACAG 401
|||||
Qy 361 atgtgtatcatggtctcctaataaacaatcacaatagatctacttacttatttatttat 420
|||||
Db 402 ATGTGTAATCATGTTCTCTAATAAATCAATCAAAATATAGTACTTACTATTGTCATTATT 461
|||||
Qy 421 tactgtttgaagctaaagacacctcacaatagatcccatccagc 464
|||||
Db 462 TACTTGTTTGAGCTAAGACCTCACAATAGAAATCCCATCCAGC 505
|||||

RESULT 6
LOCUS AW070731/c
DEFINITION IMAGE:2567361 3', mRNA sequence.
ACCESSION AW070731
VERSION AW070731.1 GI:6025729
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 414)
AUTHORS MCI-CGAP http://www.ncbi.nlm.nih.gov/mcicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP).
JOURNAL Tumor Gene Index
COMMENT Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs.r@mail.nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert length: 1011 Std Error: 0.00
Seq primer: -400P from Gibco
High quality sequence stop: 396.
FEATURES
Location/Qualifiers
1..414
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2567361"
/clone_lib="Soares_NFL_T_GBC_S1"
/note="Organ: pooled; Vector: pT7T3b-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI;
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung NBH19W, testis NHT, and B-cell
NCL-CGAP-GCB1) were mixed, and ss circles were made in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR-amplified cDNAs from pools of 5,000 clones made
from the same 3 libraries. The pools consisted of

```





```

High quality sequence stop: 379.
Location/Qualifiers
1. .393
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="IMAGE:2467736"
/tissue_type="anaplastic oligodendroglioma"
/lab_host="DH10B"
/notes="Organ: brain; Vector: pT7T3D-Pac (Pharmacia) with a
modified polylinker; Site: 1; Not 1; Site: 2; Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTCACCAATCGAGTGGAGCGGCCCATAGCTTTTIIIIIIIIIIIIIIIIIIII
T 3']; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT7T3 vector.
Library is normalized, and was constructed by Bento
Soares and M. Fatima Bonaldo."
BASE COUNT 106 a 94 c 84 g 109 t
ORIGIN

Query Match 6.3%; Score 58.4; DB 9; Length 193;
Best Local Similarity 54.9%; Pred. No. 0.0051;
Matches 113; Conservative 1; Mismatches 92; Indels 0; Gaps 0;

Oy 488 tctagtttgaaagctattataaacaacactctagtgcaattctctattcttatttgt 547
Db 361 TTTCGTTTGAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 302
Oy 548 caagtaactgggtcagacatttaccattcttcttcttcttcttcttcttcttctt 607
Db 301 CCAGGCACTGGACTGGTGGTTCATCGGATTTTCATTCAGCTTCAGGAGACCCATGA 242
Oy 508 aggaactatattatttgactacataaaatgaagaattgagacacacatttaagta 667
Db 241 AGTAGATACCAATTAGTCCCAATTTGTGGAGGAGGAGAGAGAGAGAGAGAGAGAG 182
Oy 668 attctccagggtcacacagctagaa 693
Db 181 ACTTGCCCAAGGTACACACACTAGTA 156

RESULT 11
LOCUS AQ115887
DEFINITION RPC111-56J9 JK.1 RPCI-11 Homo sapiens genomic clone RPCI-11-56J9.
DNA sequence.
ACCESSION AQ115887
VERSION AQ115887.1 GI:3492008
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 433)
AUTHORS Adams,J.M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,
Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and Venter,J.C.
Use of human BAC end Sequences for Sequence-Ready Map Building
Unpublished (1998)
Other GSSs: RPC111-56J9.TJ
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: m.adams@tigr.org
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
Research Genetics (info@resgen.com). BAC end search page:

http://www.tigr.org/tdb/hungen/bac_end_search/bac_end_search.html
Class: BAC ends.
Location/Qualifiers
1. .433
/organism="Homo sapiens"
/db_xref="GDB:7521344"
/db_xref="taxon:9606"
/clone_lib="RPCI-11-56J9"
/clone_lib="RPCI-11"
/sex="Male"
/cell_type="Lymphocytes"
/notes="Vector: pBAC3.6; Site: 1; EcoRI; Site: 2; EcoRI;
RPC111 Human Male BAC Library"
BASE COUNT 119 a 102 c 84 g 128 t
ORIGIN

Query Match 6.3%; Score 57.8; DB 12; Length 433;
Best Local Similarity 57.0%; Pred. No. 0.0069;
Matches 127; Conservative 0; Mismatches 92; Indels 4; Gaps 1;

Oy 203 gatttgaataactggatctgcacattatcagttatgtacctaagaacccgttaatt 262
Db 268 GGGTTTGAACCTGGCTCTGCCACATACCTAGCTACCTTGGGGCANGTCACGTATTC 209
Oy 263 tctcagagcctcagttctcttcttcttcttcttcttcttcttcttcttcttctt 322
Db 208 TCATCGGCCCTCAGTTCTCTCTCGAAAAATTTGGTGCAAAATAGTACATACCTGTATG 149
Oy 323 caagatctgattgacatttaataatgaagggtgaatgacagatgctgatactggtcc 378
Db 148 GTTGCTCTGGAGATTGAATTAGTTTGTAAAGAGCTTGAACAGCTGTCTGCCAGAG 89
Oy 379 aataaacatccaaatatagttacttacttacttacttacttacttacttattatt 421
Db 88 AATAAACACCAAAATATTGTTCATTACTCTTATATGATTAAATT 46

RESULT 12
LOCUS AQ598884/c
DEFINITION HS_5348_A2_H05_17A RPCI-11 Human Male BAC Library Homo sapiens
genomic clone Plate-924 Col-10 Row-O, DNA sequence.
ACCESSION AQ598884
VERSION AQ598884.1 GI:5030096
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 508)
AUTHORS Mahairas,G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
99380589
Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
or from Research Genetics (info@resgen.com). BAC end Web Server:
http://www.htsc.washington.edu
Plate: 924 row: 0 column: 10
Seq primer: 17

```

;

```

Query Match      6.1%; Score 56.8; DB 9: Length 304;
Best Local Similarity 61.7%; Pred. No. 0.012;
Matches 82; Conservative 2; Mismatches 49; Indels 0; Gaps 0;

QY 180 ttncgrtrncmgnactatntcagagtttgaaaaactgagatctgccacttctcagttatgt 239
DB 13 TTTATAGATTGAGCAATGATGCTGGGTTCAAAGTCTGGTTCCTTTCTTCTTGTTGGAATGGT 72

QY 240 gacctaaagactccgcttaattctcagagctcctcagtttcccttgctataagttggag 299
DB 73 GACITTGAGCAAGTAACTTACTTCTCTGGGGCTCACTTTCCTCCCTCTATAAATGGGAA 132

QY 300 taatattaatact 312
DB 133 TAATAATAGTACT 145

RESULT 15
A1056137/c
LOCUS A1056137 372 bp mRNA linear EST 28-AUG-1998
DEFINITION O33e08.s1 Soares_parathyroid_tumor_MbHPA Homo sapiens cDNA clone
IMAGE:1648646 3', mRNA sequence.
ACCESSION A1056137
VERSION A1056137.1 GI:3330003
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 372)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo
, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 455 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 365.
Location/Qualifiers
1..372
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1648646"
/clone_lib="Soares_parathyroid_tumor_MbHPA"
/tissue_type="parathyroid tumor"
/dev_stage="adult"
/lab_host="DH10B (ampicillin resistant)"
/notes="organ: parathyroid gland; vector: pT73D (Pharmacia
) with a modified polylinker; Site_1: Not 1; Site_2: Eco
RI; 1st strand cDNA was primed with a Not I - oligo(dT)
primer
15'-TGTTACCAATCTGAAGTGGGCGCGCCACCAATTTTTTTTTTTTTTTT
TTTT-3'. double-stranded cDNA was size selected, ligated
to Eco RI adapters (Pharmacia) digested with Not I and
cloned into the Not I and Eco RI sites of a modified pT73
vector (Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento
Soares and M. Fatima Bonaldo. RNA from sporadic parathyroid
adenomas was kindly provided by Dr. Stephen Marx, National
Institute of Diabetes and Digestive and Kidney Diseases.
NIH."
BASE COUNT 128 a 50 c 68 g 115 t
ORIGIN

```

```

Query Match      6.1%; Score 56.8; DB 9: Length 372;
Best Local Similarity 57.7%; Pred. No. 0.012;
Matches 120; Conservative 0; Mismatches 87; Indels 1; Gaps 1;

QY 224 cactatcagttatgtgaccttaaaagaactccggttaattctcagagcctcagtttccct 283
DB 334 CCCTCAGTAGTCACGTGTTCTTGGGTAAAGTCTTAACTTTTCTGTATCTCAGTTTCCTC 275

QY 284 gtcataaacttggagtaataattataactatcatcttttccaaaggattgatgtgaacatta 343
DB 274 ATCTCTAAATAGTAGGAATATTATAGTCT-ACACCTCATAGATTGTTGTGAAGATT 216

QY 344 atgaagtgaataagacagatgtgtatcatgttccctaataaacatccaaaatatagttact 403
DB 215 AACGAGTTAATAAAGTGTGTGAATGATGCTCGGCACAAAGAAATCCACTCAATTAATAGT 156

QY 404 actattgtcattattattacttctgttga 431
DB 155 AGTTCITATAGTTGTTGTATATGCTTTA 128

```

Search completed: April 29, 2002, 17:17:00  
Job time: 13219 sec







Howland, J.C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., Lehoczy, J., Lie, C., Locke, K., Macdonald, P., Marquis, N., McEvan, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrum, J., Morrow, J., Naylor, J., Norman, C.H., O'Connor, J., O'Donnell, P., Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Testaye, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X., Wyman, D., Ye, W.J., Zimmer, A. and Zody, M.

Direct Submission  
Submitted (16-NOV-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Jul 27, 2000 this sequence version replaced gi:9112670.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>  
Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: KIBR  
Web site: <http://www.seq.wi.mit.edu>  
Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
----- Project Information  
Center project name: L1907  
Center clone name: L15.N.17  
----- Summary: Statistics  
Sequencing vector: M13; M77815; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.9607.1  
Consensus quality: 152466 bases at least Q40  
Consensus quality: 154897 bases at least Q30  
Consensus quality: 156006 bases at least Q20  
Insert size: 157000; agarose-fp  
Insert size: 157295; sum-of-contigs  
Quality coverage: 5.8 in Q20 bases; agarose-fp  
Quality coverage: 5.8 in Q20 bases; sum-of-contigs

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: KIBR

Web site: <http://www.seq.wi.mit.edu>

Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

----- Project Information

Center project name: L1907

Center clone name: L15.N.17

----- Summary: Statistics

Sequencing vector: M13; M77815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.9607.1

Consensus quality: 152466 bases at least Q40

Consensus quality: 154897 bases at least Q30

Consensus quality: 156006 bases at least Q20

Insert size: 157000; agarose-fp

Insert size: 157295; sum-of-contigs

Quality coverage: 5.8 in Q20 bases; agarose-fp

Quality coverage: 5.8 in Q20 bases; sum-of-contigs

\* NOTE: This is a 'working draft' sequence. It currently consists of 10 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

\* 1 1002: contig of 1002 bp in length  
\* 1003 1102: gap of 100 bp  
\* 1103 2388: contig of 1285 bp in length  
\* 2389 2488: gap of 100 bp  
\* 2489 3525: contig of 1037 bp in length  
\* 3526 3625: gap of 100 bp  
\* 3626 16537: contig of 12902 bp in length  
\* 16538 16527: gap of 100 bp  
\* 16528 28536: contig of 12009 bp in length  
\* 28537 28736: gap of 100 bp  
\* 28737 45598: contig of 17862 bp in length  
\* 45599 45698: gap of 100 bp  
\* 46599 71912: contig of 25214 bp in length  
\* 71913 72012: gap of 100 bp  
\* 72013 96612: contig of 24600 bp in length  
\* 96613 96712: gap of 100 bp  
\* 96713 128331: contig of 31619 bp in length  
\* 128332 128431: gap of 100 bp  
\* 128432 158195: contig of 29764 bp in length.

## FEATURES

## Source

1. 158195  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="RP11-115H17"  
/clone\_lib="RPC1-11 Human Male BAC"

1. 1002

/note="assembly\_fragment"

## misc\_feature

1103..2388

## misc\_feature

2489..3525

## misc\_feature

2489..3525

/note="assembly\_fragment"  
3626..16527  
/note="assembly\_fragment"  
16628..28636  
/note="assembly\_fragment"  
clone\_end:17  
vector\_side:right  
28737..46598  
/note="assembly\_fragment"  
46699..71912  
/note="assembly\_fragment"  
72013..96612  
/note="assembly\_fragment"  
96713..128331  
/note="assembly\_fragment"  
128432..158195  
/note="assembly\_fragment"  
clone\_end:SP6  
vector\_side:left  
BASE COUNT 48628 a 30917 c 30188 g 47557 t 905 others  
OPIGIN

Query Match 83.1% Score 768.2; DB 2; Length 158195;

Best Local Similarity 92.2%; Pred. No. 4.6e-178;

Matches 850; Conservative 9; Mismatches 55; Indels 8; Gaps 6;

OY 1 caaagaaagaaagaaagcttgnaaagtagctgncatcncctcagagaccagcccttt 60

DB 14962 CCAGAGGACAGAGAAAGCTTG-AAAGTACTTGAT-CCITCTGAGGACCCGCCCTT 149019

OY 61 gcctcaaaagcaataggaatttaaaagatttncactgagaaggggncacgtttnart 120

DB 149020 GCCTCCAAAGCAATAGGAATTTAAAGATTT-CACGTGAGAAGGGGCC--ACGTTTAAG 149076

OY 121 tntnaatgntcargnarnatnccntncaaatgncrntcncactnactnrgnattgggt 180

DB 149077 TTATAAGTCTAAGGAGAAGTCATCTCAAAATGCAGCATCACATACATAGGAATTTGGG 149136

OY 181 tcccnrtncmactatnctcaggtttgaaactgagctgccactatccactgtatgtg 240

DB 149137 TTCCAGAGTCAGA--CATATCTAGGTTTGAANAATGGATCTGCCACTTATCATGTATGTG 149194

OY 241 acctaaagaactccgttaatttctcagagccctcagtttccctgtctctataaattggagct 300

DB 149195 ACCTTAAGAACTCCGTTAATTTCTCAGAGCCCTCAGTTTCTCTTAAAGTTGGAGT 149254

OY 301 aatattactactatcattttccaagagattgagtgaacattgaatgaatgaatgaatgacag 360

DB 149255 AATATTAACTATCAATTTTCCAGAGATTGATGTAACATTAAAGAGGTGAANAATGCAGAG 149314

OY 361 atgtgtatcatgatttctcctataaacaacatcccaaaatagtagtacttacttattgtcattattat 420

DB 149315 ATGTGTATCATGTTCTTAATAAACATCCAAATAATAGTACTTACTTATTGTCTATTATTAT 149374

OY 421 tacttttgaagactaaagacctcaacaatagaatccatccagcccccagagagagagc 480

DB 149375 TACTTGTITGAAGCTAAGACCTCACAATAGAAATCCATCCAGCCCCAGAGAGAGCTC 149434

OY 481 tgaatttctagtttggaagagcttataaataaacaacktctagtgcaattcttactacttg 540

DB 149435 TGAGTTTCTAGTTTGAAGAGCTAATTAATAACAGGCTAGTGTCAATTCATATTCTG 149494

OY 541 ttatgtcaagtaactgggctcagacattttacattcattgtctcttcttaagttcttaagaat 600

DB 149495 TTATGGTCAAGT-ACITGGGCTCAGCATTTTACATTCATTTCTCTTAAAGTTCTAGCAAT 149553

OY 601 gtaagcagatctactgattatttgaactacataaataaagaataatgaagctcagatataca 660

DB 149554 GTGAAGCAGGAACATATGATTATTTTACTACATNAATGAAGAAATTCAGAGCTCAGATACA 149613

OY 661 ttaagtaattctccaggggtcacacagctagaactgaaagcctgggattgattcattgata 720

DB 149613 ttaagtaattctccaggggtcacacagctagaactgaaagcctgggattgattcattgata 720

Db 149614 TTAAGTAATCTCCAGGTACACAGCTAGAACTGGCAAGCCCTGGGATTCATCCAGCA 149673  
Oy 721 tcttcagcattgaagaatcataaattgtaataaactgcaagaccctttcttcagagagc 780  
Db 149674 TCTTCAGCAATGAGAAATCATTAATGTAATACTGCAGAGCCCTTTCTTCAGAGAGC 149733  
Oy 781 tcttgctgttcacacacccactagcactgtttcttcacagggaaace ctgtgggacct 840  
Db 149734 TCTGTGCTTGCACCAAGCCCACTAGCACTTGTCTTACAGGGGAACAATGTGTGGGCT 149793  
Oy 841 ggaatcactgcacgtcacaagagtggtcttcctgaggaattattgtctctgtcagtg 900  
Db 149794 GGAATCACTGTCAGCTGCGAAGAGATGTCCTTCATGATTAATGTCCTGTCAGTGG 149853  
Oy 901 tgtgaaggcaaaaaa 922  
Db 149854 TGTAAGGCAAAAAA 149875

RESULT 2  
AL157394/c  
LOCUS  
DEFINITION Human DNA sequence from clone RP11-399019 on chr:mosome 10,  
complete sequence.  
ACCESSION AL157394  
VERSION AL157394.15 GI:15384622  
KEYWORDS HTG.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 187313)  
Blakey,S.  
Direct Submission  
Submitted (22-AUG-2001) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk  
requests: clonerequest@sanger.ac.uk  
On Aug 31, 2001 this sequence version replaced gi:14161146.  
During sequence assembly data is compared from overlapping clones.  
Where differences are found these are annotated as variations  
together with a note of the overlapping clone name. Note that the  
variation annotation may not be found in the sequence submission  
corresponding to the overlapping clone, as we submit sequences with  
only a small overlap as described above.  
This sequence was finished as follows unless otherwise noted: all  
regions were either double-stranded or sequenced with an alternate  
chemistry or covered by high quality data (i.e., phred quality >=  
30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by at least  
one plasmid subclone or more than one M13 subclone; and the  
assembly was confirmed by restriction digest. The following  
abbreviations are used to associate primary accession numbers given  
in the feature table with their source databases: Emi., EMBL; S.,  
SWISSPROT; Tr., TrEMBL; Wp., WormPEP; Information on the WormPEP  
database can be found at  
http://www.sanger.ac.uk/projects/C.elegans/wormpep  
This sequence was generated from part of bacterial clone contigs of human  
chromosome 10, constructed by the Sanger Centre Chromosome 10  
Mapping Group. Further information can be found at  
http://www.sanger.ac.uk/HGP/chr10  
RP11-399019 is from the library RP11-11.2 constructed by the group  
of Pieter de Jong. For further details see  
http://www.chori.org/bacpac/home.htm  
VCPIOR: pBACE3.5  
This sequence is the entire insert of clone RP11-399019. The true  
left end of clone RP11-496W23 is at 166408 in this sequence. The  
true right end of clone RP11-30415 is at 18704 in this sequence.  
Location/Qualifiers  
1..187313  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="10"  
/clone="RP11-399019"

FEATURES  
source

/clone\_lib="RPC1-11.2"  
100119..100156  
/note="Sequence confirmed by AC015461 sequenced by WIBR."  
100157..100198  
/note="Sequence from AC015461 sequenced by WIBR."  
105808..105972  
/note="Sequence from AC015461 sequenced by WIBR."  
105973..105989  
/note="Sequence confirmed by AC015461 sequenced by WIBR."  
BASE COUNT 55669 a 36398 c 36888 g 58358 t  
ORIGIN

Query Match 83.1%; Score 768.2; DB 9; Length 187313;  
Best Local Similarity 92.2%; Pred. No. 4.5e-178;  
Matches 850; Conservative 9; Mismatches 55; Indels 8; Gaps 4;

Oy 1 caaaggnacaggaacagcttgnaaagtactgncatnccctcgtcagggagccagcccttt 60  
Db 96976 CCAAGGACAGGACAGCTTG-AAAGTACTTGCAT-CCTTCTCGAGGACCAGCCCTTT 96919  
Oy 51 gcttcacaaagcaatagagaaatttaaaagatttncactgagaaggggncacgctllnart 120  
Db 96919 GCTTCACAAAGCAATAGGAAATTTAAAGATTT-CACTGAGAAGGGGCC--ACGTTTAA 96862  
Oy 121 tntnaatgtnrcargnanarntnccctncaaaatgncrncnctnactnactnrgnattgg 180  
Db 96861 TTATAATGTCTAAGGAGAGATCCTTCAAAATGCGAGCATCATACATAGGAATTTGG 96802  
Oy 181 tncqnrtnrcnactatntcaggtttgaaaaactggatctgcccacttatcagttatdg 240  
Db 96801 TTCAGAGTCAGA--CATATCTAGGTTTGA AAAACTGGATCTGCCACTTATCAGTTATG 96744  
Oy 241 accttaagaaacccgttaattctcagagccctcagtttcttctgctataagttggag 300  
Db 96743 ACCTTAAAGAACTCCGTTAAATTTCTCAGAGCCCTCAGTTTCTCTGTCTATAAGTTGGAG 96684  
Oy 301 aatataatactatcatcttttccaaaggattgagtgaaacatttaagtgagtgaaatgac 360  
Db 96683 AATAATATACTATCATTTTCCAGGATTGATGTGAACATTAATGAGTGAAATGACAG 96624  
Oy 361 atgtatatactggttcttaataaaacatccaaataatagtagtacttacttacttactat 420  
Db 96623 ATGTGATCATGGTTCCTCTAAATAAATCCAAATATAGTACTTACTTATGTCTATTAT 96564  
Oy 421 tactgtttgagactaaagacccctcaaatagatccatccagccaccagacagagtc 480  
Db 96563 TACTTGTITTAAGCTTAAGACCTTCAATAGAAATCCATATAGAAATCCACAGCAGAGCTC 96504  
Oy 481 tgaagtttctagtttgaagagacttataatacaacactctagtgctcaattctacttg 540  
Db 96503 TCAGTTTCTAGTTTGGAGAGAGCTATTAATAAACAACGCTCTAGTGTCAATTTCTACTTG 96444  
Oy 541 ttatggtcaagtaactgggctcagcattttacattcttcttcttaagttctagcaat 600  
Db 96443 TTATGCTCAAGT-AC TGGGCTCAGCAATTTTACATTCATTGTCTCTTTAAGTTCTAGCAAT 96385  
Oy 601 gtagacagagaaactatgattatattgactacataaataagaagaattggagctcagataca 660  
Db 96384 GTGAAGCAGGAACTATGATTATATTGACTTACATATAATGAAGAAATGAGGCTCAGATACA 96325  
Oy 661 taagtaattctccaggggttcacacagctagaaactggaagccctgggattatccatga 720  
Db 96324 TTAAGTAATTTCTCCAGGGTTCACACAGCTAGAACTGGCAAGCCCTGGGATTCATCCACA 96265  
Oy 721 tcttcagcattgaagaatcataaattgtaataaactgcaagcccttttcttcagagagac 780  
Db 96264 TCTTCAGCAATGAGAAATCATTAATGTAATACTGCAGAGCCCTTTCTTCAGAGAGC 96205  
Oy 781 tcttgctgttcacacacccactagcactgtttcttcacagggagaaactctgtggacct 840  
Db 96204 TCTTGTGCTTGCACCAAGCCCACTAGCACTTGTCTTCTACAGGGGAACATCTGTGGGCT 96145

```
Oy 841 ggaatcactgcacacagagatgttcttctgataaattatgttcttctgagtg 900
|||||
Db 96144 GGAATCAGTGCACGTCGGAAGAGATGTTCCTCAGTAAGTAAT:GTTCCTGTCAGTGG 36085
|||||

Oy 901 tgtgaaggcaaaaaa 922
|||||
Db 96084 IGTAAGGCAAAAAA 95063
|||||

RESULT 3
G43743/c
LOCUS
DEFINITION
WIAF-3334-SIS Human Thudson SANGER Homo sapiens STS genomic.
sequence tagged site.
ACCESSION
G43743
VERSION
G43743.1 GI:4192660
KEYWORDS
SIS.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 214)
Wang,D.G., Fan,J.B., Siao,C.J., Berno,A., Young,F., Sapolsky,R.,
Ghandour,G., Perkins,N., Winchester,E., Spencer,J., Kruglyak,L.,
Stein,L., Hsieh,L., Topaloglou,T., Hubbell,E., Robinson,E.,
Mittmann,M., Morris,M.S., Shen,N., Kilburn,D., Rioux,J.,
Nusbaum,C., Rozen,S., Hudson,T.J., Lipshutz,R., Chee,M. and
Lander,E.S.
Large-scale identification, mapping, and genotyping of
single-nucleotide polymorphisms in the human genome
Science 280 (5365), 1077-1082 (1998)
98248615
9582121
Synonyms: stSG15752a, stSG15752
Contact: Thomas Hudson
Whitehead Institute/MIT Center for Genome Research
Whitehead Institute for Biomedical Research
9 Cambridge Center, Cambridge MA 02142 USA
Tel: 617 252 1900
Fax: 617 252 1902
Email: thudson@genome.wi.mit.edu
Primer A: CATCAGAGCAACATCTCTTCG
Primer B: ATTCTCCAGGGTCACACG
STS size: 214
PCR Profile:
  Presoak: 94 degrees C for 4.00 minutes
  Denaturation: 94 degrees C for 50.0 seconds
  Annealing: 58 degrees C for 1.50 minutes
  Polymerization: 72 degrees C for 1.00 minutes
  PCR Cycles: 30
  Thermal Cycler: custom built by IAS, Costar, Cambridge MA

Protocol:
  Template: 10 ng
  Primer: each 5 pM
  dNTPs: 4 mM
  Taq Polymerase: 0.5 U
  Total Vol: 20 uL

Buffer:
  Mg2+: 1.5 mM
  KCl: 50 mM
  Tris-HCl: 10 mM
  Gelatin: .001 %
  Location/Qualifiers
    1..214
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /map="112.70 cR from top of Chr10 linkage group"
      /clone_lib="Human Thudson SANGER"
      /notes="human STS created from EST in the Sanger database"
      1..214
      primer_bind
      primer_bind
        complement(195..214)

FEATURES
  source
```

```
BASE COUNT 52 a 48 c 56 g 57 t 1 others
ORIGIN
Query Match 19.9% Score 184.2; DB 11; Length 214;
Best Local Similarity 97.2%; Pred. No. 9,1e-35;
Matches 208; Conservative 0; Mismatches 4; Indels 2; Gaps 2;

Oy 568 attctccagggtcacacagtagaactggcaag-cctgggattgacatattctcc 726
|||||
Db 214 ATTCTCCAGGGTCACACAGCTAGAAACTGCAAGCCCTGGGATTGATCATGATC 155
|||||

Oy 727 agcattgaagaatcataatgtaataactgcaaggcctttctctcagaagact-cctg 785
|||||
Db 154 AGCATTTGAAGAAATCATAAATGTAATACATGCAAGGCTTTTCTCTCAGAGAG 95
|||||

Oy 786 gtgcttgaccacacccactagcactgttctctacagggggaacatctgtgggctc 845
|||||
Db 94 GTGCTTGCAACCAACCCACTAGCACTGTCTCTACAGGGGAACATCTGTGGCC 35
|||||

Oy 845 tcaatgcacgtgcgaagagatgttcttctgagt 879
|||||
Db 34 TCACCTGCACGTCGCAAGAGATGTTCCTTCTGATG 1
|||||

RESULT 4
G43744/c
LOCUS
DEFINITION
WIAF-3335-SIS Human Thudson SANGER Homo sapiens STS genomic.
sequence tagged site.
ACCESSION
G43744
VERSION
G43744.1 GI:4192661
KEYWORDS
SIS.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 214)
Wang,D.G., Fan,J.B., Siao,C.J., Berno,A., Young,P., Sapolsky,R.,
Ghandour,G., Perkins,N., Winchester,E., Spencer,J., Kruglyak,L.,
Stein,L., Hsieh,L., Topaloglou,T., Hubbell,E., Robinson,E.,
Mittmann,M., Morris,M.S., Shen,N., Kilburn,D., Rioux,J.,
Nusbaum,C., Rozen,S., Hudson,T.J., Lipshutz,R., Chee,M. and
Lander,E.S.
Large-scale identification, mapping, and genotyping of
single-nucleotide polymorphisms in the human genome
Science 280 (5366), 1077-1082 (1998)
98248615
9582121
Synonyms: stSG15752b, stSG15752
Contact: Thomas Hudson
Whitehead Institute/MIT Center for Genome Research
Whitehead Institute for Biomedical Research
9 Cambridge Center, Cambridge MA 02142 USA
Tel: 617 252 1900
Fax: 617 252 1902
Email: thudson@genome.wi.mit.edu
Primer A: CATCAGAGCAACATCTCTTCG
Primer B: ATTCTCCAGGGTCACACG
STS size: 214
PCR Profile:
  Presoak: 94 degrees C for 4.00 minutes
  Denaturation: 94 degrees C for 50.0 seconds
  Annealing: 58 degrees C for 1.50 minutes
  Polymerization: 72 degrees C for 1.00 minutes
  PCR Cycles: 30
  Thermal Cycler: custom built by IAS, Costar, Cambridge MA

Protocol:
  Template: 10 ng
  Primer: each 5 pM
  dNTPs: 4 mM
  Taq Polymerase: 0.5 U
  Total Vol: 20 uL

TITLE
JOURNAL
MEDLINE
PUBMED
COMMENT
```



```

* 16453 15552: gap of 100 bp
* 16553 17254: contig of 702 bp in length
* 17255 17354: gap of 100 bp
* 17355 18030: contig of 676 bp in length
* 18031 18130: gap of 100 bp
* 18131 18844: contig of 714 bp in length
* 18845 18944: gap of 100 bp
* 18945 19558: contig of 714 bp in length
* 19559 19758: gap of 100 bp
* 19759 20457: contig of 699 bp in length
* 20458 20557: gap of 100 bp
* 20558 21263: contig of 706 bp in length
* 21264 21363: gap of 100 bp
* 21364 22070: contig of 707 bp in length
* 22071 22170: gap of 100 bp
* 22171 22879: contig of 709 bp in length
* 22880 22979: gap of 100 bp
* 22880 23688: contig of 709 bp in length
* 23689 23788: gap of 100 bp
* 23789 24498: contig of 710 bp in length
* 24499 24598: gap of 100 bp
* 24599 25321: contig of 723 bp in length
* 25322 25421: gap of 100 bp
* 25422 26114: contig of 693 bp in length
* 26115 26214: gap of 100 bp
* 26215 26917: contig of 703 bp in length
* 26918 27017: gap of 100 bp
* 27018 27720: contig of 703 bp in length
* 27721 27820: gap of 100 bp
* 27821 28477: contig of 657 bp in length
* 28478 28577: gap of 100 bp
* 28578 29281: contig of 704 bp in length
* 29282 29381: gap of 100 bp
* 29382 30091: contig of 710 bp in length
* 30092 30191: gap of 100 bp
* 30192 30901: contig of 710 bp in length
* 30902 31001: gap of 100 bp
* 31002 31694: contig of 693 bp in length
* 31695 31794: gap of 100 bp
* 31795 32493: contig of 699 bp in length
* 32494 32593: gap of 100 bp
* 32594 33299: contig of 706 bp in length
* 33300 33499: gap of 100 bp
* 33400 34081: contig of 682 bp in length
* 34082 34181: gap of 100 bp
* 34182 34898: contig of 707 bp in length
* 34899 34988: gap of 100 bp
* 34989 35675: contig of 688 bp in length
* 35677 35776: gap of 100 bp
* 35777 36487: contig of 711 bp in length
* 36488 36587: gap of 100 bp
* 36588 37285: contig of 698 bp in length
* 37286 37385: gap of 100 bp
* 37386 38079: contig of 694 bp in length
* 38080 38179: gap of 100 bp
* 38180 38852: contig of 673 bp in length
* 38853 38952: gap of 100 bp
* 38953 39667: contig of 715 bp in length
* 39668 39767: gap of 100 bp
* 39768 40483: contig of 716 bp in length
* 40484 40583: gap of 100 bp
* 40584 41278: contig of 655 bp in length
* 41279 41378: gap of 100 bp
* 41379 42080: contig of 702 bp in length
* 42081 42180: gap of 100 bp
* 42181 42896: contig of 715 bp in length
* 42897 42996: gap of 100 bp
* 42997 43693: contig of 697 bp in length
* 43694 43793: gap of 100 bp
* 43794 44459: contig of 706 bp in length
* 44500 44599: gap of 100 bp
* 44600 45319: contig of 720 bp in length
* 45320 45419: gap of 100 bp

```

```

* 45420 46130: contig of 711 bp in length
* 46131 46230: gap of 100 bp
* 46231 46934: contig of 704 bp in length
* 46935 47034: gap of 100 bp
* 47035 47743: contig of 709 bp in length
* 47744 47843: gap of 100 bp
* 47844 48554: contig of 711 bp in length
* 48555 48654: gap of 100 bp
* 48555 49364: contig of 710 bp in length
* 49365 49464: gap of 100 bp
* 49465 50142: contig of 678 bp in length
* 50143 50242: gap of 100 bp
* 50243 50951: contig of 709 bp in length
* 50952 51051: gap of 100 bp
* 51052 51762: contig of 711 bp in length
* 51763 51862: gap of 100 bp
* 51863 52553: contig of 691 bp in length
* 52554 52653: gap of 100 bp
* 52654 53360: contig of 707 bp in length
* 53361 53460: gap of 100 bp
* 53461 54142: contig of 682 bp in length
* 54143 54242: gap of 100 bp
* 54243 54952: contig of 710 bp in length
* 54953 55052: gap of 100 bp

Query Watch          9.0%: Score 83.6; DB 2; Length 66933;
Best Local Similarity 58.2%; Pred. No. 2.1e-10;
Matches 255; Conservative 0; Mismatches 151; Indels 32; Gaps 5:

OY 277 ttctctgtctataagttgggaataataataactatcatcttttccaaggaattgatgtg 335
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 35433 TTTTITTTTNNTTTTTGGTtAGGAATATCAACACCATTATTTTCAAGGAGATAGATG 35374

OY 337 aacattaaa-----tgaggtaaatgacagatgtgtatcattgggttcctaataaacatcc 399
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 35373 TAGGCTAAAGTGTATGCCAAAATTAATGCAAGTATATGGGGTGTGTATGATAATAATCC 35314

OY 390 aaaaatagtacttactatgttcatttacttacttctgttgaagctcaagacctcacaat 449
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 35313 AACAAAGAGTAAACAAGACTGTCACTACTGT---CTGTGTGAATTTCAAGACCATACACC 35257

OY 450 aaaaatcccatccagccaccagacagagytctgagtttcttagtttgggaagagctattaa 509
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 35256 TAGGGAAACGTCAGTAGAACGACTAGAGA-----GTTCTGATGTGGATGAGCCGATAA 35204

OY 510 ataacaactctagtctcaattctatcttatttacttctgttgaagtaactgggctcagcattt 569
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 35203 ATGATAGCATCTACTGTCTAGTTCCTCAGAGGCCCTGGTCAAGCACAGGGTGT----- 35150

OY 570 tacattcattgtctctttaaagttctagcaatgtgaagcaggaaactatgattatattgact 629
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 35151 -----TGCAAGTGCACCTAACTCCTAATGATTTAAGGCAGGAACCT---ATTATCCTAACT 35099

OY 630 acataaatgaagaattgaggtcagatcacatcacattaaatctccagaggtcacacagct 689
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 35098 ACCCTAGTGAATGAATTAAGGCTCAGATGCACAAAGTATTTTTCCTAAAGTCACATAGCT 35039

OY 690 aqaactggcgaagcctgg 707
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 35038 AAGCGTGGCAATGCCCTAG 35021

```

```

RESULT 5
AP0000851/c
LOCUS
DEFINITION
  AP0000851
  Homo sapiens genomic DNA, chromosome 11q, clone:RP11-817J15,
  complete sequences.
ACCESSION
  AP0000851
  GI:13488770
VERSION
  HTG.
SOURCE
  Homo sapiens
  ORGANISM
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

URL: http://hgp.gsc.riken.go.jp/, Tel: 81-42-778-9923,  
Fax: 81-42-778-9924  
On May 31, 2000 this sequence version replaced gi:6997739.  
\*\*\*\*\* Genome Center  
Center: RIKEN Genomic Sciences Center (GSC)  
Center code: RIKEN  
Web site: http://hgp.gsc.riken.go.jp/  
Contact: hattori@hgp.gsc.riken.go.jp  
\*\*\*\*\* Project Information  
Center project name: HumDraft11  
Center clone name: RP11-750P5  
\*\*\*\*\* Summary Statistics  
Sequencing vector: PCR products; 100% of reads  
Chemistry: Dye-terminator ET-amersham; 100% of reads  
Assembly program: Phrap; version 0.990329  
Consensus quality: 149299 bases at least Q40  
Consensus quality: 161098 bases at least Q30  
Consensus quality: 167078 bases at least Q20  
Insert size: 169959; sum-of-contigs  
Quality coverage: 4.01x in Q20 bases; sum-of-contigs  
\*\*\*\*\*  
NOTE: This is a 'working draft' sequence. It currently consists of 35 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved  
1 15727 contig of 15727 bp in length  
15828 27185 contig of 11359 bp in length  
27287 37387 contig of 10101 bp in length  
37488 47288 contig of 9801 bp in length  
47389 55992 contig of 8604 bp in length  
56093 63632 contig of 7540 bp in length  
63733 73241 contig of 9509 bp in length  
73342 80659 contig of 7318 bp in length  
80760 89061 contig of 8302 bp in length  
89162 94975 contig of 5814 bp in length  
95076 100459 contig of 5744 bp in length  
100560 106303 contig of 5421 bp in length  
106404 111824 contig of 5674 bp in length  
111925 117598 contig of 5674 bp in length  
117699 123774 contig of 6076 bp in length  
123875 128021 contig of 4147 bp in length  
128122 134542 contig of 6421 bp in length  
134643 138972 contig of 4330 bp in length  
139073 143438 contig of 4265 bp in length  
143438 146720 contig of 3283 bp in length  
146821 149011 contig of 2191 bp in length  
149112 151132 contig of 2021 bp in length  
151233 154403 contig of 3171 bp in length  
154504 156567 contig of 2064 bp in length  
156668 158701 contig of 2034 bp in length  
158802 160732 contig of 1931 bp in length  
160833 162522 contig of 1690 bp in length  
162623 164124 contig of 1502 bp in length  
164225 165574 contig of 1350 bp in length  
165675 167245 contig of 1571 bp in length  
167346 168587 contig of 1242 bp in length  
168688 169757 contig of 1070 bp in length  
169858 170997 contig of 1140 bp in length  
171098 172148 contig of 1051 bp in length  
17249 173359 contig of 1111 bp in length  
Sequence updated (26-May-2000).  
\* NOTE: This is a 'working draft' sequence. It currently consists of 35 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.  
1 15727: contig of 15727 bp in length

URL: http://hgp.gsc.riken.go.jp/, Tel: 81-42-778-9923,  
Fax: 81-42-778-9924  
On May 31, 2000 this sequence version replaced gi:6997739.  
\*\*\*\*\* Genome Center  
Center: RIKEN Genomic Sciences Center (GSC)  
Center code: RIKEN  
Web site: http://hgp.gsc.riken.go.jp/  
Contact: hattori@hgp.gsc.riken.go.jp  
\*\*\*\*\* Project Information  
Center project name: HumDraft11  
Center clone name: RP11-750P5  
\*\*\*\*\* Summary Statistics  
Sequencing vector: PCR products; 100% of reads  
Chemistry: Dye-terminator ET-amersham; 100% of reads  
Assembly program: Phrap; version 0.990329  
Consensus quality: 149299 bases at least Q40  
Consensus quality: 161098 bases at least Q30  
Consensus quality: 167078 bases at least Q20  
Insert size: 169959; sum-of-contigs  
Quality coverage: 4.01x in Q20 bases; sum-of-contigs  
\*\*\*\*\*  
NOTE: This is a 'working draft' sequence. It currently consists of 35 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved  
1 15727 contig of 15727 bp in length  
15828 27185 contig of 11359 bp in length  
27287 37387 contig of 10101 bp in length  
37488 47288 contig of 9801 bp in length  
47389 55992 contig of 8604 bp in length  
56093 63632 contig of 7540 bp in length  
63733 73241 contig of 9509 bp in length  
73342 80659 contig of 7318 bp in length  
80760 89061 contig of 8302 bp in length  
89162 94975 contig of 5814 bp in length  
95076 100459 contig of 5744 bp in length  
100560 106303 contig of 5421 bp in length  
106404 111824 contig of 5674 bp in length  
111925 117598 contig of 5674 bp in length  
117699 123774 contig of 6076 bp in length  
123875 128021 contig of 4147 bp in length  
128122 134542 contig of 6421 bp in length  
134643 138972 contig of 4330 bp in length  
139073 143438 contig of 4265 bp in length  
143438 146720 contig of 3283 bp in length  
146821 149011 contig of 2191 bp in length  
149112 151132 contig of 2021 bp in length  
151233 154403 contig of 3171 bp in length  
154504 156567 contig of 2064 bp in length  
156668 158701 contig of 2034 bp in length  
158802 160732 contig of 1931 bp in length  
160833 162522 contig of 1690 bp in length  
162623 164124 contig of 1502 bp in length  
164225 165574 contig of 1350 bp in length  
165675 167245 contig of 1571 bp in length  
167346 168587 contig of 1242 bp in length  
168688 169757 contig of 1070 bp in length  
169858 170997 contig of 1140 bp in length  
171098 172148 contig of 1051 bp in length  
17249 173359 contig of 1111 bp in length  
Sequence updated (26-May-2000).  
\* NOTE: This is a 'working draft' sequence. It currently consists of 35 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.  
1 15727: contig of 15727 bp in length

URL: http://hgp.gsc.riken.go.jp/, Tel: 81-42-778-9923,  
Fax: 81-42-778-9924  
On May 31, 2000 this sequence version replaced gi:6997739.  
\*\*\*\*\* Genome Center  
Center: RIKEN Genomic Sciences Center (GSC)  
Center code: RIKEN  
Web site: http://hgp.gsc.riken.go.jp/  
Contact: hattori@hgp.gsc.riken.go.jp  
\*\*\*\*\* Project Information  
Center project name: HumDraft11  
Center clone name: RP11-750P5  
\*\*\*\*\* Summary Statistics  
Sequencing vector: PCR products; 100% of reads  
Chemistry: Dye-terminator ET-amersham; 100% of reads  
Assembly program: Phrap; version 0.990329  
Consensus quality: 149299 bases at least Q40  
Consensus quality: 161098 bases at least Q30  
Consensus quality: 167078 bases at least Q20  
Insert size: 169959; sum-of-contigs  
Quality coverage: 4.01x in Q20 bases; sum-of-contigs  
\*\*\*\*\*  
NOTE: This is a 'working draft' sequence. It currently consists of 35 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved  
1 15727 contig of 15727 bp in length  
15828 27185 contig of 11359 bp in length  
27287 37387 contig of 10101 bp in length  
37488 47288 contig of 9801 bp in length  
47389 55992 contig of 8604 bp in length  
56093 63632 contig of 7540 bp in length  
63733 73241 contig of 9509 bp in length  
73342 80659 contig of 7318 bp in length  
80760 89061 contig of 8302 bp in length  
89162 94975 contig of 5814 bp in length  
95076 100459 contig of 5744 bp in length  
100560 106303 contig of 5421 bp in length  
106404 111824 contig of 5674 bp in length  
111925 117598 contig of 5674 bp in length  
117699 123774 contig of 6076 bp in length  
123875 128021 contig of 4147 bp in length  
128122 134542 contig of 6421 bp in length  
134643 138972 contig of 4330 bp in length  
139073 143438 contig of 4265 bp in length  
143438 146720 contig of 3283 bp in length  
146821 149011 contig of 2191 bp in length  
149112 151132 contig of 2021 bp in length  
151233 154403 contig of 3171 bp in length  
154504 156567 contig of 2064 bp in length  
156668 158701 contig of 2034 bp in length  
158802 160732 contig of 1931 bp in length  
160833 162522 contig of 1690 bp in length  
162623 164124 contig of 1502 bp in length  
164225 165574 contig of 1350 bp in length  
165675 167245 contig of 1571 bp in length  
167346 168587 contig of 1242 bp in length  
168688 169757 contig of 1070 bp in length





Choepel, Y., Colangelo, M., Collins, S., Collamore, A., Cooke, F.,  
D'Arellano, K., Dewar, K., Diaz, J.S., Dodge, S., Ferreira, P.,  
FitzHugh, W., Gage, D., Galagan, J., Gardina, S., Ginde, S., Gorsette, M.,  
Graham, L., Grand-pierre, N., Hagos, B., Heaford, A., Horton, L.,  
Jillev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., LePocque, K.,  
Lmazares, R., Landers, T., Lehoczy, J., Levine, R., Lieu, C., Liu, G.,  
Macdonald, P., Marquis, N., McCarthy, M., McEwan, P., McKernan, K.,  
McPheeters, R., Meldrum, J., Menous, L., Mihov, T., Mlenga, V.,  
Morrow, J., Murphy, T., Naylor, J., Norman, C.H., O'Connor, T.,  
O'Donnell, P., O'Neill, D., Oliver, T.M., Oliver, J., Peterson, K.,  
Pierre, N., Pisan, C., Pollara, V., Raymond, C., Pieback, M., Riley, R.,  
Pogor, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Sever, P.,  
Sougnuez, C., Spencer, B., Stange-Thomann, N., Stojanovic, N.,  
Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,  
Tirrell, A., Travers, M., Trigglio, J., Vassiliou, H., Viel, R., Vo, A.,  
Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, S., Zainoun, J.,  
Zimmer, A. and Zody, M.

Direct Submission  
Submitted (13-AUG-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Nov 22, 2000 this sequence version replaced gi:975795.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RH/RepeatMasker.html>

----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: <http://www-seq.wi.mit.edu>  
Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
----- Project Information  
Center project name: L10752  
Center clone name: 151\_M13  
----- Summary Statistics  
Sequencing vector: Plasmid; n/a; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.96071  
Consensus quality: 177020 bases at least Q49  
Consensus quality: 181299 bases at least Q50  
Consensus quality: 182918 bases at least Q20  
Insert size: 182000; agarose-1p  
Insert size: 184023; sum-of-contigs  
Quality coverage: 5.2 in 020 bases; agarose-1p  
Quality coverage: 5.2 in 020 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently  
consists of 14 contigs. The true order of the pieces  
is not known and their order in this sequence record is  
arbitrary. Gaps between the contigs are represented as  
runs of N, but the exact sizes of the gaps are unknown.  
This record will be updated with the finished sequence  
as soon as it is available and the accession number will  
be preserved.

1 2204: contig of 2204 bp in length  
2205 2304: gap of 100 bp  
2305 5276: contig of 3972 bp in length  
5277 6376: gap of 100 bp  
6377 9572: contig of 3195 bp in length  
9573 9672: gap of 100 bp  
9673 13976: contig of 4304 bp in length  
13977 14076: gap of 100 bp  
14077 18878: contig of 4802 bp in length  
18879 18978: gap of 100 bp  
18979 24619: contig of 5641 bp in length  
24620 24719: gap of 100 bp  
24720 34230: contig of 9511 bp in length  
34231 34330: gap of 100 bp  
34331 41754: contig of 7424 bp in length  
41755 41854: gap of 100 bp  
41855 53704: contig of 11850 bp in length  
53705 53804: gap of 100 bp  
53805 67811: contig of 14007 bp in length  
67812 67911: gap of 100 bp  
67912 84130: contig of 16219 bp in length

84131 84230: gap of 100 bp  
84231 95885: contig of 11355 bp in length  
95886 95885: gap of 100 bp  
95886 151552: contig of 55867 bp in length  
151553 151652: gap of 100 bp  
151653 185323: contig of 33671 bp in length.

FEATURES  
Location/Qualifiers  
source  
1..185323  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="11"  
/map="11"  
/clone\_lib="RPCI-11 Human Male BAC"  
1..2204  
/note="assembly\_fragment"  
clone\_end:SP6  
vector\_side:left  
2305..6276  
/note="assembly\_fragment"  
6377..9572  
/note="assembly\_fragment"  
9673..13976  
/note="assembly\_fragment"  
14077..18878  
/note="assembly\_fragment"  
18979..24619  
/note="assembly\_fragment"  
24720..34230  
/note="assembly\_fragment"  
34331..41754  
/note="assembly\_fragment"  
41855..53704  
/note="assembly\_fragment"  
53805..67811  
/note="assembly\_fragment"  
67912..84130  
/note="assembly\_fragment"  
84231..95885  
/note="assembly\_fragment"  
95886..151552  
/note="assembly\_fragment"  
151653..185323  
/note="assembly\_fragment"  
clone\_end:T7  
vector\_side:right

BASE COUNT 54890 a 36117 c 36604 g 56410 t 1302 others

Query Match 7.6% Score 70.2; DB 2: Length 185323;  
Best Local Similarity 62.8%; Pred. No. 3.6e-07;  
Matches 125; Conservative 0; Mismatches 73; Indels 1; Gaps 1:

Qy 494 ttggagagacttataataacacactctagtcactctacactctgtatggtcaagta 553  
|||||  
Db 20652 TTCGTGTTATGATGATATACATATATACATCATCTTATTCAGTACTTTCAGGC 20711  
|||||  
Qy 554 actggactngcattttacattctctcttttaagttctagcaatggaagcaggaac 613  
|||||  
Db 20712 ATGTGCCAAGCA-TTACACACACTTCTCATTTAAITACACACTTGATACAGTAGG 20770  
|||||  
Qy 514 tatgattatattgactacataataataagagctcaggtccagatataagtaattctc 673  
|||||  
Db 20771 TATTAATATCTTCACCTTCATTAATGAAGAACTAAAGTTAGAAAGTTAATACCTTAC 20830  
|||||  
Qy 574 ccagggtacacagctaga 692  
|||||  
Db 20831 CCAAGATCAACACAGCTATA 20849  
|||||

RES:11 9  
HS57:19

**LOCUS** HS57G9 113872 bp DNA linear PFI 12-DEC-1999  
**DEFINITION** Human DNA sequence from clone CTA-57G9 on chromosome 22q12.1. Contains part of the gene for a novel Kringle and CUB domain protein, ESI5, STSS, CA repeat polymorphisms, GSSs and a CpG Island, complete sequence.  
**ACCESSION** Z95116  
**VERSION** Z95116.1 GI:2832592  
**KEYWORDS** HTG: ca repeat polymorphism; CpG island; CUB domain; Kringle domain; human.  
**SOURCE** Homo sapiens  
**ORGANISM** Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
**REFERENCE** 1 (bases 1 to 113872)  
**AUTHORS** McLaren,S.  
**TITLE** Direct Submission  
**JOURNAL** Submitted (09-DEC-1999) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk  
**COMMENT** requests: clonerequest@sanger.ac.uk  
 On Feb 5, 1998 this sequence version replaced gi:2578126.  
 During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.  
 This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.  
 The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WormPEP; Information on the WormPEP database can be found at [http://www.sanger.ac.uk/projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/projects/C_elegans/wormpep) This sequence was generated from part of bacterial clone contigs of human chromosome 22, constructed by the Sanger Centre Chromosome 22 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr22>  
 CTA-57G9 is from the human BAC library described in U-J. Kim et al. (1995) Genomics 34, 213-218.  
**FEATURES** SOURCE  
 Location/Qualifiers  
 1..113872  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="22"  
 /map="q12.1"  
 /clone="CTA-57G9"  
 /clone\_lib="CIT978SK-A1"  
 complement(1)..284)  
 /note="match: GSS: Em:B14184  
 match: STS: Em:B14184"  
 35..392  
 /note="match: GSS: Em:B14225"  
 complement(305..682)  
 /note="match: GSS: Em:B14096"  
 593..618  
 /note="13 copies 2 mer at 96 conserved"  
 740..1035  
 /note="AluY repeat: matches 1..296 of consensus"  
 1326..1841  
 /note="MLT1E repeat: matches 44..549 of consensus"  
 1857..2202  
 /note="MLT1A2 repeat: matches 24..374 of consensus"  
 2346..2519  
 /note="MER21B repeat: matches 160..318 of consensus"  
 2520..2665  
 /note="FLAM\_C repeat: matches 1..133 of consensus"

2666..3075  
 /note="MER21B repeat: matches 318..792 of consensus"  
 3241..3458  
 /note="MLT1D repeat: matches 1..242 of consensus"  
 4048..4345  
 /note="MLT1D repeat: matches 206..505 of consensus"  
 4665..4996  
 /note="match: GSS: Em:A0119749"  
 join(4811..4935,8717..8870,20796..21128,22078..22209)  
 /gene="BK57G9.1"  
 join(4811..4935,8717..8870,20796..21128,22078..22209)  
 /gene="BK57G9.1"  
 /note="supported by GENSCAN  
 match: ESTs: Em:AA506741 Em:AA670682  
 match: Proteins: Tr:P91972 Sw:P98065 Tr:O15113 Tr:O351113  
 Tr:Q24132 Tr:O70244 Tr:O57433 Tr:O88204 Sw:P97435  
 Tr:O60494 Tr:O57658 Tr:O75074 Tr:O57434 Tr:O08628  
 Tr:O61398 Sw:P98066 Tr:O23995 Sw:P97333 Tr:O08859  
 Tr:O95917"  
 /codon\_start=3  
 /evidence="not experimental"  
 /product="BK57G9.1 (novel Kringle and CUB domain protein)"  
 /protein\_id="CA662952.1"  
 /db\_xref="GI:6572252"  
 /translation="PCNLGCKYKDHGNPPPLTGTSTKNTLTQTCISFCRSQRFKAG  
 MESGVACFCGNPNPDYATGRVCYWTIRVGASHIHFSPLFDIROSADVLLDGVTH  
 SAMSVVVSYPDFDIATGRVCYWTIRVGASHIHFSPLFDIROSADVLLDGVTH  
 RVLARFHGRSRRPLSFNVSLDFVILYFESDRINQAGFAVLQYQVKEELPOERPAVNC  
 TVAEVITEQANLSVSAARSKVLYVITIS"  
 4957..5312  
 /note="LIMB6 repeat: matches 5808..6175 of consensus"  
 5309..6111  
 /note="LIM1 repeat: matches 5016..5828 of consensus"  
 6112..6422  
 /note="AluSg repeat: matches 1..308 of consensus"  
 6423..6546  
 /note="LIM1 repeat: matches 5828..5952 of consensus"  
 6502..6965  
 /note="LIMD2 repeat: matches 4987..5472 of consensus"  
 6966..7272  
 /note="AluSx repeat: matches 3..311 of consensus"  
 7273..7944  
 /note="LIMD2 repeat: matches 5472..6069 of consensus"  
 7945..8246  
 /note="AluSx repeat: matches 1..302 of consensus"  
 8247..8328  
 /note="LIMD2 repeat: matches 6069..6153 of consensus"  
 8368..8426  
 /note="MIR repeat: matches 201..262 of consensus"  
 8959..9050  
 /note="L2 repeat: matches 2641..2732 of consensus"  
 9708..10119  
 /note="MLT1C repeat: matches 8..464 of consensus"  
 10556..10724  
 /note="MIR repeat: matches 74..252 of consensus"  
 10783..10914  
 /note="LIMB6 repeat: matches 5448..5581 of consensus"  
 10915..11122  
 /note="AluSg/x repeat: matches 81..296 of consensus"  
 11123..11661  
 /note="LIMB6 repeat: matches 5576..6122 of consensus"  
 11710..11811  
 /note="MER47 repeat: matches 2224..2323 of consensus"  
 12111..12286  
 /note="MIR repeat: matches 47..262 of consensus"  
 12663..12810  
 /note="AluSg/x repeat: matches 141..292 of consensus"  
 13522..13690  
 /note="MIR repeat: matches 78..252 of consensus"  
 13684..13741  
 /note="L2 repeat: matches 2608..2666 of consensus"  
 14535..14774  
 /note="MIR repeat: matches 18..259 of consensus"



```
* be preserved.
* 1 67: contig of 67 bp in length
* 68 167: gap of 100 bp
* 168 1872: contig of 1705 bp in length
* 1873 1972: gap of 100 bp
* 1973 7959: contig of 5987 bp in length
* 7960 8055: gap of 100 bp
* 8060 14082: contig of 6023 bp in length
* 14083 14182: gap of 100 bp
* 14183 24347: contig of 10165 bp in length
* 24348 24447: gap of 100 bp
* 24448 44018: contig of 19571 bp in length
* 44019 44118: gap of 100 bp
* 44119 52009: contig of 17891 bp in length
* 52010 52103: gap of 100 bp
* 52110 84737: contig of 22628 bp in length
* 84738 84837: gap of 100 bp
* 84838 115856: contig of 31019 bp in length
* 115857 115957: gap of 100 bp
* 115957 162051: contig of 46105 bp in length.
FEATURES
    source
        Location/Qualifiers
            1..162051
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /chromosome="3"
                /map="3"
                /clone_lib="RPC1-11 Human Male BAC"
                /clone="RPC1-11"
            1..67
                /note="assembly_fragment"
                clone_end:17
                vector_side:right
            158..1872
                /note="assembly_fragment"
            1973..7959
                /note="assembly_fragment"
            8060..14082
                /note="assembly_fragment"
                clone_end:56
                vector_side:left
            14183..24347
                /note="assembly_fragment"
            24448..44018
                /note="assembly_fragment"
            44119..52009
                /note="assembly_fragment"
            52110..84737
                /note="assembly_fragment"
            84838..115856
                /note="assembly_fragment"
            115957..162051
                /note="assembly_fragment"
BASE COUNT 50903 a 30373 c 31811 g 48073 t 901 others
ORIGIN
Query Match 7.3%; Score 67.4; DB 2: Length 162051;
Best Local Similarity 56.4%; Pred. No. 1.8e-06;
Matches 145; Conservative 1; Mismatches 107; Indels 5; Gaps 1:
Oy 502 gctattaatacaacckctagctgcaattctctattgctgctgaagtaactgcctt 561
Db 47111 GTTACTAACCCACTAATGCTACATTTATTAACTATTATTATGCTAGGCACGTCTCT 47170
Oy 562 cagcattttacattcattctcttattttagttctagca-----tctgaagcaggaaactat 616
Db 47171 TAGCACTTGAGATAAGCACTTTCATTCATTAATCTATATAGGAAGCAGCTAT 47230
Oy 517 gattattatcactacataaataaataatgagggctcagatcatatgaattctccca 676
Db 47231 TATTATCAATCATTTTCAAAATAAGAAATAAGGCACAGAGTAAAGTAAATTGCCCA 47250
Oy 577 ggggtcacacagctagaaactggcaagcctgggattgattccatgatcttccacattgaag 736
```

```
Db 47291 AGGTCACACAGCTGTGAAGTGACAGAGAGGAGGATTTAGCCAGGCACTCTGGTTTCAGA 47350
Oy 737 aatcataaatgtataataac 755
Db 47351 GCATACATACCATTAATTACC 47369
RESULT 11
LOCUS AL133551 175940 bp DNA linear PRI 21-AUG-2000
DEFINITION Human DNA sequence from clone RP11-57G10 on chromosome 10 contains a J-domain containing protein (JDP1) isoform B, the SIRT1 gene (Sirt-like proteins (sirtuins) type 1), part of a novel gene similar to KIAA0032, two ribosomal pseudogenes, 2 CpG islands, ESTs, STSS and USSs, complete sequence.
ACCESSION AL133551
VERSION 1
KEYWORDS HIT: J-domain; RPL12; RPL21; SIRT1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 175940)
AUTHORS Wilson,S.
TITLE Direct Submission
JOURNAL Submitted (21-AUG-2000) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Jul 23, 2000 this sequence version replaced gi:9367364.
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.
The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em., EMBL; Sw., SWISSPROT; Tr., TrEMBL; Wp., WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/Celegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 10, constructed by the Sanger Centre Chromosome 10 Mapping Group. Further information can be found at http://www.sanger.ac.uk/HGP/Chr10
RP11-57G10 is from the library RPC1-11.1 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see http://bacpac.med.buffalo.edu/
VECTOR: pBACE3.6
This sequence is the entire insert of clone RP11-57G10 The true left end of clone RP11-474D14 is at 172652 in this sequence.
FEATURES
    source
        Location/Qualifiers
            1..175940
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /chromosome="10"
                /clone="RP11-57G10"
                /clone_lib="RPC1-11.1"
            1..2318
                /note="LIPAL0 repeat: matches 3815. .6158 of consensus"
            2338..2767
                /note="LIMA4 repeat: matches 4946. .5385 of consensus"
            2768..3058
                /note="LIMA4 repeat: matches 1. .296 of consensus"
            3059..3371
                /note="LIMA4 repeat: matches 5385. .5704 of consensus"
            3372..3682
                /note="LIMA4 repeat: matches 5385. .5704 of consensus"
```





Qy 507 taataacaackcttagtgcataactgtttaaggccaagtaactgggctcagca 566  
| | | | | | | | | | | | | | | | | | | | | | | |  
Db 12394 TCACCACCACTATCTAATGTTGATTGGCCCTTGC AATGTGCCAGCTTCTGACTAGCT 12453  
  
Qy 567 tttaacattcatctcttaagtctcaga-----atgtgaagcagaacataatt 620  
| | | | | | | | | | | | | | | | | | | | | | | |  
Db 12454 CTTTGCACTGGATTCATNTTAATCTTCACMACCAGTATTTGAAGTAGACTATATT 12513  
  
Qy 621 atattgactacataaaatgaagaattgagctcgagacatactaatcccccaggt 680  
| | | | | | | | | | | | | | | | | | | | | | | |  
Db 12514 ATCTTCATTTATAGAGGAGGAACTGAGGTTTAGAGAGGTTAAATAATTTTGCACAAGT 12573  
  
Qy 681 cacacagtagaacctgccaaagctggagtgatccatgatcttcacagcattgaagaatc 740  
| | | | | | | | | | | | | | | | | | | | | | | |  
Db 12574 CATAGCTAGTAGTACTGACTGGGATTTGAATTTAAACCTGGCGAGCTGACTCCA 12633  
  
Qy 741 ataataataataactcgaagccttttctccagaagaagctccctgggctt 791  
| | | | | | | | | | | | | | | | | | | | | | | |  
Db 12634 AGCCAGCTACTTAAGCTGTCGACAGTACTTCCAAAGGCGTCTGTCAGCTT 12684

RESULT 13

AC078778 linear HTG 25-JUN-2001

LOCUS Homo sapiens chromosome 12 clone RP11-968A15, WORKING DRAFT

DEFINITION SEQUENCE, 8 unordered pieces.

ACCESSION AC078778

VERSION GI:14547475

KEYWORDS HTG; HIGS\_PHASEI; HTGS\_DRAFT; HTGS\_FULLTOP.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:  
Mammalia: Eutheria: Primates: Catarrhini: Hominoidea: Homo.  
1 (bases 1 to 207856)

REFERENCE Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,  
Alsbrooks,S.L., Amaralunga,H.C., Are,J.R., Banks,T., Barbaria,J.,  
Benton,J., Birmage,K., Blankenburg,K., Bonnin,D., Bouck,J.,  
Bowles,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,  
Hurch,P., Burket,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Chen,  
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,  
Chen,Z., Chowdhury,I., Christopoulos,C., Cleveland,C.D., Cox,C.,  
Coyle-McDoll,L., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,  
Day-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,  
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,  
Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,  
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,  
Foster,P., Franz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,  
Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,  
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,  
Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,  
Hollins,B., Homs,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,  
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivest,S.,  
Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,  
Kovar,C., Kratochiv,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,  
Lewis,L., Li,J., Li,Z., Licharge,O., Lieu,C., Liu,J., Liu,W.,  
Loulseghe,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,  
Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,  
Martinez,E., Massey,E., Mawhinney,E., McLeod,M.P., Meador,M.,  
Mel,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,  
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,  
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokkenwo,S.,  
Oguh,M., Okwuonu,G., Orangunye,N., Oviedo,R., Pace,A., Payton,B.,  
Peery,J., Perez-Ll., Peters,L., Pickens,R., Primus,E., Pu,L.L.,  
Quilly,M., Ren,Y., Rives,M., Rojas,A., Rojokoban,I., Rolfe,M.,  
Rut,S., Savory,G., Scherer,S., Scott,G., Shen,H., Shoohtari,N.,  
Stinson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,  
Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamiera,A., Tamerisa,K.,  
Tanaka,H., Tansey,J., Taylor,C., Taylor,T., Terfrod,B., Thomas,N.,  
Tiana,H., Tansley,J., Taylor,C., Taylor,T., Terfrod,B., Thomas,N.,

Thomas, S., Osmani, K., Vagstad, B., Verastegui, R., Warren, R., Washington, C., Wall, R., Wang, S., Ward-Moore, S., Williams, A., Wleczyk, R., Wooden, S., Watlington, S., Williams, G., Williamson, A., Wolcott, M., Woodson, D., Worley, K., Wu, C., Wu, Y., Wu, J., Zhou, J., Zorrilla, S., Nelson, D.





Matches	102:	Conservative	0:	Mismatches	51:	Indels	0:	Gaps	0:
QY	529	attctatctattatgatcgaactgagctcaacattttacatttcatttctctttaa	588						
Db	54958	ATTGAGCACTGTTATGTTCTAGACATTTGTTTCAACAACTTTTCCATGCAATTAGCTCAITTA	54899						
QY	589	agtcttagcaatgtgaacgagaactatgatttatattgactacataaatgaagaataatga	548						
Db	54898	ATCACAAACACCTCAGTAGTGAGTACTACTATTATTCCTCATTTTAAAGATAGGAAACTGA	54839						
QY	649	ggctcagatcacatgaatgaattctcccgagggtcacacagctag	621						
Db	54838	GGTACTGAACAGTTAACTACCTTGCCTGGGTACATGGCTAG	51795						
RESULT 15	AF391809/c	AF391809	24771 bp	DNA	linear	FBI 20-JUL-2001			
LOCUS		Homo sapiens coagulation factor II (thrombin) receptor (F2R) gene							
DEFINITION		complete cds.							
ACCESSION		AF391809							
VERSION		AF391809.2	GI:14971463						
KEYWORDS									
SOURCE		human.							
ORGANISM		Homo sapiens							
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.							
AUTHORS		1 (bases 1 to 24771)							
TITLE		Rieder,M.J., Carrington,D.P., Chung,M.-W., Lee,K.L., Foel,C.L., Yi,O. and Nickerson,D.A.							
JOURNAL		Direct Submission							
REFERENCE		Submitted (15-JUN-2001) Molecular Biotechnology, University of Washington, 1705 NE Pacific, Seattle, WA 98195, USA							
AUTHORS		2 (bases 1 to 24771)							
JOURNAL		Rieder,M.J. and Nickerson,D.A.							
REMARK		Direct Submission							
COMMENT		Submitted (20-JUL-2001) Molecular Biotechnology, University of Washington, 1705 NE Pacific, Seattle, WA 98195, USA							
FEATURES		Sequence update by submitter							
source		On Jul 20, 2001 this sequence version replaced oi:14583134.							
repeat_region		To cite this work please use: SeattleSNPs. NHLBI Program for Genomic Applications, UW-FHCRC, Seattle, WA (URL: http://pga.mbt.washington.edu).							
variation		Location/Qualifiers							
variation		1..24771							
variation		/organism="Homo sapiens"							
variation		/db_xref="taxon:9606"							
variation		55..211							
variation		/rpt_family="MER104"							
variation		/rpt_type="dispersed							
variation		234							
variation		/frequency="0.07"							
variation		/replace="C"							
variation		352							
variation		/frequency="0.42"							
variation		/replace="A"							
variation		535							
variation		/frequency="0.01"							
variation		/replace="A"							
variation		709							
variation		/frequency="0.02"							
variation		/replace="T"							
variation		811..1058							
variation		/rpt_family="AluJb"							
variation		/rpt_type="dispersed							
variation		1184							
variation		/frequency="0.01"							
variation		/replace="G"							
variation		1531..1543							
variation		/frequency="0.54"							
variation		/replace="A"							
variation		1545..2053							
variation		/note="Region not scanned for variation"							

mrna	join(1807. .2237,18156. .21315) /gene="F2R" /product="coagulation factor II (thrombin) receptor" 1807. .21315 /gene="F2R" 2081 /gene="F2R" /frequency="0.01" /replace="A" join(2150. .2237,18156. .19345) /gene="F2R" /codon_start=1 /product="coagulation factor II (thrombin) receptor" /protein_id="AAK69768.1" /db_xref="GI:14583135" /translation="MGPRRLLVAAFCSLGCLLSARTARRPESKATNATLDRSFL LPNNDKYEPWEDEEKNEGLTEYRLVLSINKSPLOKLPAFISEDAGYLTSWLT LFVSVYGVFVSVLPLNMAIVFLLKMKVKPAVYMLHLATADVLVSVLPFKLS YIFSGSDWQFGECLRFYTAFFYCNMVASILLMTVISIDRFNAVYPMOSLSWRTLGR ASFTCLAIWALAIAGVPLLLKEQTIOVPLNITITCHDVLNETLLLEGYYAYYSFASFA VFFPPLIISTVCYVSIIRCLSSAVANRKKSLFSAAFVCFIICFQPINLLILI AHVFSLSHTSITEAMAFYLLICVGVSSICIDPLIYYASSECQRYVYSTLCKKES DPSSVNSSQGLMASKMDTCSSNLNNSIYKKLT"
variation	2372 /gene="F2R" /frequency="0.05" /replace="A" 2502 /gene="F2R" /frequency="0.01" /replace="A" 2762 /gene="F2R" /frequency="0.06" /replace="T" 2860 /gene="F2R" /frequency="0.40" /replace="A" 2930 /gene="F2R" /frequency="0.22" /replace="C" 2940 /gene="F2R" /frequency="0.01" /replace="G" 2952 /gene="F2R" /frequency="0.01" /replace="T"
variation	3252 /gene="F2R" /frequency="0.01" /replace="C" 3481 /gene="F2R" /frequency="0.39" /replace="C"
variation	3525 /gene="F2R" /frequency="0.10" /replace="T" 3941. .4047 /rpt_family="MLR" /rpt_type="dispersed"
variation	3948 /gene="F2R" /frequency="0.21" /replace="G" 4017 /gene="F2R" /frequency="0.24"
repeat_region	
variation	
variation	

```

/replace="C"
4038
/genes="F2R"
/frequency="0.03"
/replace="A"
4087
/genes="F2R"
/frequency="0.02"
/replace="T"
4107
/genes="F2R"
/frequency="0.03"
/replace="T"
4336
/genes="F2R"
/frequency="0.41"
/replace="A"
4393
/genes="F2R"
/frequency="0.03"
/replace="A"
4517
/genes="F2R"
/frequency="0.07"
/replace="T"
4925
/genes="F2R"
/frequency="0.01"
/replace="G"
4971..5125
/rpt_family="MIR"
/rpt_type=dispersed
5072
/genes="F2R"
/frequency="0.01"
/replace="C"
5128
/genes="F2R"
/frequency="0.22"
/replace="G"
5150..5441
/rpt_family="AluSx"
/rpt_type=dispersed
5159
/genes="F2R"
/frequency="0.02"
/replace="T"
5210
/genes="F2R"
/frequency="0.01"
/replace="C"
5276
/genes="F2R"
/frequency="0.03"
/replace="A"
5428
/genes="F2R"
/frequency="0.01"
/replace="A"
5472
/genes="F2R"
/frequency="0.40"
/replace="A"
5480
/genes="F2R"
/frequency="0.01"
/replace="A"
5515
/genes="F2R"
/frequency="0.01"
/replace="A"

```

Query Match

7.0% Score 64.6: DB 9: Length 24771:

```

Best Local Similarity 56.8%: Pred. No. 1.le-05:
Matches 138: Conservative 0: Mismatches 104: Indels 1: Gaps 1:
Oy 556 tgggtcagcattttacattcattgtctctttaagttctagcaatgtgaagcagaacta 615
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 9390 TGTGCCAGATACTTCACITGCCTTAGTTCATATATAATTCTTAACATTCTATGACATGAGCA 9331
Oy 616 tgattatagtacataaaatgaagaattgaggctcagatacatttaagtaattctccc 675
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 9330 TTATAACTACCATITTTATAATGACAACATGGAGGCACAGAGGCTAAGTAACATAATCC 9271
Oy 675 agggtcacacagcta-gaactggcgaagcctgggattgacctgatcttccagcattga 734
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 9270 AAAGTCACACAGCTATGAAGTAGAGGAGCCAGGGTTTGAACCTAGGTCTGCCAGCTCCA 9211
Oy 735 agaatacataaatgtataaataactgcaaggccttttctctcagaagagctcctggttga 794
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 9210 AAGGTCAIACCTTTTGACCACCTCCACGCTATTCTCTCCTTTAAAACTTGTGATTGAC 9151
Oy 795 cca 797
    |||
Db 9150 CTA 9148

```

Search completed: April 29, 2002, 19:50:29  
 Job time: 22428 sec



```

Db 362 TTCATGCTCTCTGGATTTCATCCAAATTAATNACCT 400
      BE186013
      LOCUS      316 bp mRNA linear EST 22-JUN-2000
      DEFINITION CMI-HT0738-250400-199-bll HT0738 Homo sapiens cDNA, mRNA sequence.
      ACCESSION BE186013
      VERSION BE186013.1 GI:8665197
      KEYWORDS EST.
      SOURCE human.
      ORGANISM Homo sapiens
      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
      Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
      1 (bases 1 to 316)
      Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
      Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bata,G.S., Simpson,D.H.,
      Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
      ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
      Simpson,A.J.
      Shotgun sequencing of the human transcriptome with ORF expressed
      sequence tags
      Proc. Natl. Acad. Sci. U.S.A. 97 (7). 3491-3496 (2000)
      20202663
      Contact: Simpson A.J.G.
      Laboratory of Cancer Genetics
      Ludwig Institute for Cancer Research
      Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
      Brazil
      Tel: +55-11-2704922
      Fax: +55-11-2707001
      Email: asimpson@ludwig.org.br
      This sequence was derived from the FAPESP/LICR Human Cancer Genome
      Project. This entry can be seen in the following URL
      (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=6t2-CMI-HT0738-250
      400-199-bll&t3=2000-04-25&t4=1)
      Seq primer: puc 18 forward
      High quality sequence start: 4
      High quality sequence stop: 316.
      Location/Qualifiers
        1..316
          /organism="Homo sapiens"
          /db_xref="taxon:9606"
          /clone_lib="HT0738"
          /dev_stage="Adult"
          /note="Organ: Head/neck; Vector: puc18; Site_1: SmaI;
          Site_2: SmaI; A mini-library was made by cloning products
          derived from ORESTES PCR (U.S. Letters Patent application
          No. 196,716 - Ludwig Institute for Cancer Research)
          profiles into the pUC 18 vector. Reverse transcription of
          tissue mRNA and cDNA amplification were performed under
          low stringency conditions."
          BASE COUNT 69 a 93 c 71 g 82 t 1 others
      ORIGIN
      Query Match 61.2%; Score 293.6; DB 9; Length 316;
      Best Local Similarity 98.4%; Pred. No. 6.6e-40;
      Matches 307; Conservative 0; Mismatches 4; Indels 1; Gaps 1;
      QY 60 agcagtcctgtagtactctggagatttccatcttcttggtctc-tgccagaatccgacaa 118
      DB 5 AGCAGCTCTGACTCTGGGAGTTCCTATCTTCCGGTCTCGTATCAGATCCGACAA 64
      QY 119 cagctgtccagctgacacgtatccagctactggtctgctgatgatgaagccctgatg 178
      DB 65 CAGCTGCTCAGTGCACATATCCAGTACGGGTCTGCTGATGATGAGAGCCCTGATG 124
      QY 179 ctgaaccactgctgtrgaaccactgcaccactgctgctctactac-aactgcaccaccg 238
      DB 125 CTGNAACCACTGCTGCTGCAACCACTGCGACCACTGCTGCTCTCTACCACTGCAACCCG 184

```

---

```

QY 239 ctgcttccaccactgctgtaagacattccagttttaccacaatgggttggaatctcc 298
      DB 185 CTGCTTCTACCAGCTGCTGTAAGACATTCAGTTTACCCAAATGGGTGGGATCTCC 244
      QY 299 cgaatggttagagtggtgctccctgaatgaatcagcttgaagctctctcgaattggtcaca 358
      DB 245 CGAATGGTAGAGTGTGTCCCTGAGATGAATCAGCTTGAGTCTTCTGCAATTTGTCACAA 304
      QY 359 ctattcatgctt 370
      DB 305 CTATTCATGCTT 316

RESULT 15
BE925938
LOCUS      316 bp mRNA linear EST 02-OCT-2000
DEFINITION CMI-BT0738-240800-199-bll BT0738 Homo sapiens cDNA, mRNA sequence.
ACCESSION BE925938
VERSION BE925938.1 GI:10453014
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 316)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bata,G.S., Simpson,D.H.,
Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7). 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=6t2-CMI-BT0738-240
800-199-bll&t3=2000-08-24&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 4
High quality sequence stop: 316.
Location/Qualifiers
  1..316
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone_lib="BT0738"
    /dev_stage="Adult"
    /note="Organ: breast; Vector: puc18; Site_1: SmaI; Site_2:
    SmaI; A mini-library was made by cloning products derived
    from ORESTES PCR (U.S. Letters Patent application No. 196
    716 - Ludwig Institute for Cancer Research) profiles
    into the pUC 18 vector. Reverse transcription of tissue
    mRNA and cDNA amplification were performed under low
    stringency conditions."
    BASE COUNT 69 a 93 c 71 g 82 t 1 others
  ORIGIN
  Query Match 61.2%; Score 293.6; DB 10; Length 316;
  Best Local Similarity 98.4%; Pred. No. 6.6e-40;
  Matches 307; Conservative 0; Mismatches 4; Indels 1; Gaps 1;
  QY 50 agcagtcctgtagtactctggagatttccatcttcttggtctc-tgccagaatccgacaa 118

```











```

Db 133 GCTGCTCCAGCTGACACGATCCAGCTACTGGTCTCTGCTGATGATGAAGCCCTGATGTT 192
Oy 181 gaaccactgtgtgaaaccactggaaccactgtgtgtctctaccactgcaaccaccgt 240
Db 193 GAAACACCTGTGCTGCAACCTGCGACCACTGCTGCTCTCTACCACTGCAACCAACCGCT 252
Oy 241 gctctaccactgtctgaaagacattccagttttaccaccaatgggtggggtatctccg 300
Db 253 GCTTCTACCACTGCTGTAAGACATTCAGTTTTACCAAAATGGGTGGGATCTCCCG 312
Oy 301 aatgtaagaagtgtctccctgagatggaatcagcttgagttctctcgaattgtgcacaact 360
Db 313 AATGGTAGAGTGTGCTGAGATGAATCAGCTTGAGTCTTCGCAATGGTCACAACT 372
Oy 361 attcatgctctctgattccatccaaactacttacccttgctcctacagata 408
Db 373 ATTATGCTTCTGATTTTCAATCCAACTACTTACCTTGGCTACACTA 420

```

```

RESULT 5
AI139456/c 402 bp mRNA linear EST 27-OCT-1998
LOCUS qc20g01.x1 Soares fetal heart NbHL19W Homo sapiens cDNA clone
DEFINITION IMAGE:1710192 3' similar to contains element MFR40 repetitive
element ;, mRNA sequence.
ACCESSION AI139456
VERSION AI139455.1 GI:3645428
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 402)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: rcp@pshs.fda.gov
This clone is available royalty-free through LLNL: contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 510 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 398.
Location/Qualifiers
1..402
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Soares_fetal_heart_NbHL19W"
/sex="unknown"
/dev_stage="19 weeks"
/lab_host="DH10B (ampicillin resistant)"
/notes="Organ: heart; Vector: pT73D (Pharmacia) with a
modified polylinker; Site:1: Not I; Site:2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAAGTGGCGCGCGCACTTTTITTTTTTTTTT 3'].
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT73 vector
(Pharmacia). Library went through one round of
normalization to a Cot - 5. Library constructed by:
M.Fatima Bonaldo. This library was constructed from the
same fetus as the fetal lung library, Soares fetal lung
NbHL19W."

```

```

BASE COUNT 115 a 115 g 100 t
ORIGIN

Query Match 92.84; Score 397.2; DB 9; Length 402;
Best Local Similarity 99.34; Pred. No. 2.7e-57;
Matches 399; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

Oy 67 ctggactcttggagatttccatcttcttggctctctgcccagaatccgacaacagctgct 126
Db 402 CTGGTACTCTTGGAGATTCCATCTATCTGGTCTCTGCCCCAATCCGACACACAGCTGCT 343
Oy 127 ccagctgacacgtatccagctactgctgctgctgctgctgctgctgctgctgctgctgct 186
Db 342 CCAGCTGACACGATCCAGCTACTGGTCTGCTGATTAATGAAGCCCTGATGCTGAAACC 283
Oy 187 actgctgctgcaaccctcgcgaccactgctgctctctaccactgcaaccacgctgctctct 246
Db 282 ACTGCTGCTGCAACCACTGCGACCACTGCTGCTCCAAACCACTGCAACCAACCGCTGCTCT 223
Oy 247 accactgctctaaacacattccagttttaccaccaatgggtggggtatcccgaaatggt 306
Db 222 ACCACTGCTCTGTAAGACATTCAGTTTTACCAAAATGGTGGGATCTCCCGAATGCT 153
Oy 307 agagtgtgctccctgagatggaatcagcttgagctctctgcaattggtcacaacttgcatt 366
Db 152 AGAGTGTGCTGAGATGAATCAGCTTGAGTCTTCTGCAATGGTGCACAACTATTCAT 103
Oy 367 gctctctgattctccaaactacttacccttgctcctacgataccccctttatctcctaact 426
Db 102 GCCTTCTGCTGATTTTCAATCCAACTACTTACCTTGGCTAGGATATCCCTTTATCTTAATC 43
Oy 427 agtttatcttcttcaataaaataaataaactatgagcaacaaa 468
Db 42 AGTTATTTCTTTCAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 1

RESULT 6
BG202312 464 bp mRNA linear EST 21-APR-2001
LOCUS RST121568 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.
DEFINITION BG202312
ACCESSION BG202312
VERSION BG202312.1 GI:13723999
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 464)
AUTHORS Harrington,J.J., Sherf,B., Rundlett,S., Jackson,P.D., Perry,R.,
Cain,S., Leventhal,C., Thornton,M., Ramachandran,R., Whittinton,J.,
Lerner,L., Costanzo,D., McElligott,K., Booser,S., Mays,R., Smith
,E., Veloso,N., Klika,A., Hess,J., Cothren,K., Lo,K., Offenbacher
,J., Danzig,J. and Ducar,M.
Creation of genome-wide protein expression libraries using random
activation of gene expression
Nat. Biotechnol. 19 (5), 440-445 (2001)
JOURNAL 2127151
MEDLINE
COMMENT Contact: Scott J. Cain
Athersys, Inc.
3201 Carnegie Ave, Cleveland, OH 44115, USA
Tel: 216 431 9900
Fax: 216 361 9596
Email: scain@athersys.com
High quality sequence stop: 464.
Location/Qualifiers
1..464
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Athersys RAGE Library"
/cell_line="HT1080"
/notes="See 'Creation of Genome-wide Protein Expression
Libraries using Random Activation of Gene Expression',
Nature Biotechnology, in press. Note that even though the
cell type indicated is HT1080, since a random activation
method was used, these sequence tags are not necessarily
expressed in HT1080 under normal circumstances."

```

```

BASE COUNT 126 a 119 c 91 g 128 t
ORIGIN

```



BASE	COUNT	ORIGIN
92 a	128 c	81 g 134 t 1 others

/note="Organ: heart; Vector: pT73D (Pharmacia) with a modified polylinker; Site\_1: Not 1; Site\_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer 15'-GTGACCAATCTGAAGTCGAGCGCGCACTTTTITTTTTTTT 3', 1GTCACCTATCTGAAGTCGAGCGCGCACTTTTITTTTTTTT 3', RI double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cleaved into the Not I and Eco RI sites of a modified pT73 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by M.Rafaina Bonaldo. This library was constructed from the same fetus as the fetal lung library, Soares fetal lung NbuH19W."

Query Match	88.0%	Score 422.4:	DB 10:	Length 436:
Best Local Similarity	99.5%:	Prod. No. 1.1e-61:		
Matches 434:	Conservative 0:	Mismatches 1:	Indels 1:	Gaps 1:
<hr/>				
Qy	10	gcatttttgctgctccctgatcttcagggtcacacaccatgaagttctttagcagtccttg	69	
Dbb	1	GCATTTTGTGTCGCCCTGATCTTCATGTCCACCACCAATGAAGTTCITTAGCAGTCCITG	50	
<hr/>				
Qy	70	gtactctggaggatttcacattcttcttggtctctgccagaatacccacaacagctgtcca	129	
Dbb	61	GTACTCTGGGAGITTCCAATTCTTGCTGCTGCCAGAAICCGACAACAGTGTCCA	120	
<hr/>				
Qy	130	gctgacacagtaccagctactggtctcgctgcatgatatgaagccccctaatgctgaaacacct	189	
Dbb	121	GCTGACACGTATCCAGCTACTTGGTCTGTCGTGATGATGAAGCCCCCTGATGCTGAACCACT	180	
<hr/>				
Qy	190	gtgtgtgcaaaccactgagaccactgtgtgtctctaccactgcaaccaccgctgttttacc	249	

OY 250 actgctctaaagacattccagttttacccaaatggattggagatctccggaatgtaga 309  
 |||||  
 Db 241 ACTGCTCTGAAGACATTCCAGTTTTTACCNAATGGGTGGGATCTCCGAATGGTAGA 300  
 OY 310 gtgtgtccctggagatggaaatcagcttgagttcttctgcaatt-ggtcacaaactattcatgc 368  
 |||||  
 Db 301 GTGTGTCTCGAGATGGGAATCAGCTTGAGTCTTCTGCAATTTNGTCAACAATATTCTATGC 360  
 OY 369 ttccctgtaattcatcccaactacttacctgtgctctacatatcccccttatctcttaalcaag 428  
 |||||  
 Db 361 TTCCGTGATTTATCCCAACTACTTACCTTGGCTACGAATATCCCCITTATCTCTAATCAAS 420  
 OY 429 ttattttcttttcaaa 444  
 |||||  
 Db 421 TTATATTTCTTTCAAA 436

RESULT	2
AW070916/c	
LOCUS	AW070916          420 bp        mRNA       linear    EST-20-OCT-2000
DEFINITION	xaij001.xl NCI_CGAP_Brl8 Homo sapiens cDNA clone IMAGE:256385.3. similar to contains element TAR1 repetitive element ; , mRNA sequence.
ACCESSION	AW070916
VERSION	AW070916
KEYWORDS	AW070916.1 GI:5025914
SOURCE	EST.
ORGANISM	human.
	Homo sapiens
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE	1 (bases 1 to 420)
AUTHORS	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL	Unpublished (1997)
COMMENT	Contact: Robert Strausberg, Ph.D. Email: coapbs@mail.nih.gov

```

Tissue Procurement: Chris Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: Life
Technologies, Inc. cDNA Library Arrayed by: Christa Prange, The
I.M.A.G.E. Consortium DNA Sequencing by: Washington University
Genome Sequencing Center
Clone distribution: NCI-CCAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/obrp/image/image.html
Insert length: 545 Std Error: 0.00
Seq primer: -400P from Gibco.

FEATURES
    source
        1..420
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="IMAGE:2568385"
            /clone_lib="NCI-CCAP_Brl8"
            /tissue_type="four pooled high-grade tumors, including two
            primary tumors and two metastatic to ovary"
            /lab_host="DH10B"
            /note=Organ: breast; Vector: pCMV-SPORT6; Site_1: SalI;
            Site_2: NotI; Cloned unidirectionally. Primer: Oligo dr.
            Library constructed by Life Technologies."
            119 a 76 c 116 g 109 t

```

	Query Match	87.2%	Score 418.4	DB 9	Length 420
	Best Local Similarity	99.8%	Pred. No. 7.5e-61		
	Matches 419	Conservative 0	Mismatches 1	Indels 0	Gaps 0
Qy	59	taccagctctgtactcttgggaattccatctttctgggtctctgcccagaatccgacaa	118		
Db	420	TAGCAGTCCCTGGTACTCTTGGAGATTCCACTTCTTGGTCTGTGCCAGAAATCCGACAA	361		
Qy	119	cagctgtctcagctgcacqctatccagctactggtctgctgatgatgaagccctgatg	178		
Db	350	CAGCTGCTCCAGCTGACACGTATCCAGCTACTGTGTCTGTGATGATCAGGCCCTGATG	301		
Qy	179	ctgaaccactgtctgtgcacaccactgcgaccactgctctctaccactgcgaaccacg	238		
Db	300	CTGAACCACTGCTGTGCACACACTGCGACCACTGCTGCTTACCACCTGCACACACCG	241		

[illegible]

RESULT	3
W72838/c	
LOCUS	W72838 432 bp mRNA linear EST 16-OCT-1996
DEFINITION	d57g12.sl Soares_fetal_heart_NbH19W Homo sapiens cDNA clone IMAGE:344806 3' similar to contains element MER40 repetitive element ;, mRNA sequence.
ACCESSION	W72838
VERSION	W72838.1 GI:1362814
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 432)

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 29, 2002, 17:16:35 ; Search time 6783.7 seconds  
(without alignments)  
955.015 Million cell updates/sec

Title: US-09-248-178-56  
Perfect score: 480  
Sequence: 1 ggccttggaagcattttgtc.....gcaacaaaaaaaaaaaaa 480

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues  
Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST.\*

- 1: em\_estba.\*
- 2: em\_esthum.\*
- 3: em\_estin.\*
- 4: em\_estnu.\*
- 5: em\_estov.\*
- 6: em\_estpl.\*
- 7: em\_estro.\*
- 8: em\_hic.\*
- 9: gb\_estl.\*
- 10: gb\_est2.\*
- 11: gb\_hic.\*
- 12: gb\_gss.\*
- 13: em\_gss\_hum.\*
- 14: em\_gss\_inv.\*
- 15: em\_gss\_pln.\*
- 16: em\_gss\_vrt.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	422.4	88.0	436	10 W72837	W72837 zd57g12.r1
C 2	418.4	87.2	420	9 AW070016	AW070016 xat1d01.x
C 3	418.4	87.2	432	10 W72838	W72838 zd57g12.s1
C 4	403.2	84.0	423	10 BF742809	BF742809 I12-BT080
C 5	397.2	82.8	402	9 A1139456	A1139456 qc20g01.x
6	372.8	77.7	464	10 BG202312	BG202312 RST21668
7	358.6	74.7	376	10 BG218084	BG218084 RST37810
8	337.4	70.3	391	10 BG185269	BG185269 RST4208 A
9	314	65.4	314	12 AZ694036	AZ694036 AST-2HBBG
C 10	311.8	65.0	343	10 BG202313	BG202313 RST21669
C 11	306.6	63.9	345	9 BE032421	BE032421 CM1-BT073
C 12	306.6	63.9	313	9 AW176044	AW176044 QV0-BT010
C 13	302	62.9	403	10 R75793	R75793 y121f09.r1
14	293.6	61.2	316	9 BE186013	BE186013 CM1-HT073
15	293.6	61.2	316	10 BE926938	BE926938 CM1-BT073
16	291	60.6	294	9 AA340069	AA340069 EST45219
C 17	259	54.0	443	9 A1127172	A1127172 qb99a08.x

5/2/02

18	251.2	52.3	308	9	A1905687	A1905687 IL-BT095-
C 19	224.8	46.7	462	9	AW294149	AW294149 UI-H-B12-
C 20	224	46.7	224	9	AW291950	AW291950 UI-H-B12-
C 21	219.4	45.7	251	10	BG197186	BG197186 RST16424
C 22	204.4	42.6	221	9	BE074512	BE074512 IL5-BT057
C 23	200.2	41.7	230	10	BE815819	BE815819 PM3-BN017
C 24	189.4	39.5	248	10	BE815824	BE815824 PM3-BN017
C 25	177	36.9	211	10	BG183176	BG183176 RST2047 A
C 26	176	36.7	212	10	BG192597	BG192597 RST11712
C 27	165.4	34.7	211	10	BG207535	BG207535 RST27017
C 28	161	33.5	193	9	A1905624	A1905624 CM-BT094-
C 29	156	32.5	224	9	A1905837	A1905837 IL-BT099-
C 30	150.4	31.3	178	9	A1905623	A1905623 CM-BT094-
C 31	133	27.7	136	9	A1905633	A1905633 CM-BT094-
C 32	106.6	22.2	498	10	B1279448	B1279448 UI-R-DA0-
C 33	105.8	22.0	492	10	B1279431	B1279431 UI-R-DA0-
C 34	103.8	21.6	224	10	BE815853	BE815853 PM3-BN017
C 35	103.4	21.5	486	10	B1282803	B1282803 UI-R-DA0-
C 36	103.4	21.5	490	10	B1279159	B1279159 UI-R-DA0-
C 37	103.4	21.5	490	10	B1279279	B1279279 UI-R-DA0-
C 38	103.4	21.5	493	10	B1279513	B1279513 UI-R-DA0-
C 39	103.4	21.5	493	10	B1279518	B1279518 UI-R-DA0-
C 40	103.4	21.5	495	10	B1279214	B1279214 UI-R-DA0-
C 41	103.4	21.5	497	10	B1279243	B1279243 UI-R-DA0-
C 42	103	21.5	487	10	B1279415	B1279415 UI-R-DA0-
C 43	101.8	21.2	499	10	B1279473	B1279473 UI-R-DA0-
C 44	101.8	21.2	500	10	B1279161	B1279161 UI-R-DA0-
C 45	101.8	21.2	503	10	B1279551	B1279551 UI-R-DA0-

## ALIGNMENTS

RESULT 1  
W72837 436 bp mRNA linear EST 16-OCT-1996  
LOCUS zd57g12.r1 Soares\_fetal\_heart NBHH19W Homo sapiens cDNA clone  
DEFINITION IMAGE:344806 5' similar to contains element MER40 repetitive  
element 1, mRNA sequence.

ACCESSION W72837  
VERSION W72837.1 GI:1382813

KEYWORDS EST.  
SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 436)  
AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Riffe, L., Rohlfsing, T., Soares, M., Tan, F., Trevaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.

TITLE The WashU-Merck EST Project

JOURNAL Unpublished (1995)

COMMENT Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
This clone is available royalty-free through LNL: contact the  
INHERT Consortium (info@image.llnl.gov) for further information.  
Insert Length: 508 Std Error: 0.00  
Seq primer: mob.REGA-ET  
Location/Qualifiers  
1. 436  
/organism="Homo sapiens"  
/db\_xref="GDB:1270181"  
/db\_xref="taxon:9606"  
/clone="IMAGE:344806"  
/clone\_lib="Soares\_fetal\_heart\_NBHH19W"  
/sex="unknown"  
/dev\_stage="19 weeks"  
/lab\_host="DH10B (ampicillin resistant)"





VERSION	AE003526.2	GI:10727878
KEYWORDS	HTG.	
SOURCE	fruit fly.	
ORGANISM	Drosophila melanogaster	
	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.	
REFERENCE	1 (bases 1 to 280887)	
AUTHORS	Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F., George,R.A., Lewis,S.A., Richards,S., Ashburner,M., Henderson,S.N., Sutton,G.G., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X., Brannon,R.C., Rogers,Y.H., Blazej,R.G., Champe,M., Pfeiffer,B.D., Wan,K.H., Doyle,C., Baxter,E.G., Heit,G., Nelson,C.R., Gabor Miklos,G.L., Abril,J.F., Agbayani,A., An,H.J., Andrews,Pfannkuch,C., Balgwin,D., Ballex,R.M., Basu,A., Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y., Benos,P.V., Berman,B.P., Bhandari,D., Bolshakov,S., Borkota,D., Botchan,M.P., Bouck,J., Brokstein,P., Brottier,P., Burtis,K.C., Piusan,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I., Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dex,I., Dietz,S.M., Podoson,K., Delcher,A., Dong,M., Dugan-Rocha,S., Dunkov,B.C., Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Fertileta,S., Fleischmann,W., Foster,C., Gabrielian,A.E., Garg,N.S., Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z., Guan,P., Harris,M., Harris,N.L., Harvey,D., Heiman,T.J., Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J., Wai,M.H., Ibeagwam,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z., Kennison,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C., Kravitz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A., Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Mattei,B., McIntosh,I.C., McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Moharry,C., Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L., Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K., Pao,S., Russkern,D.P., Pacieb,J.M., Palazzolo,M., Pittman,G.S., Pan,S., Pollard,J., Puri,V., Reese,M.G., Reinert,K., Remington,K., Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Kiamos,I., Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C., Stapleton,M., Strong,R., Sun,E., Svitskas,R., Tector,C., Turner,R., Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wassarman,D.A., Weinstein,G.M., Weissbach,J., Williams,S.M., Woodage,T., Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Ye,R.F., Zaveri,J.S., Zhan,M., Zhang,G., Zhao,O., Zheng,L., Zheng,X.H., Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O., Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.	
	The genome sequence of Drosophila melanogaster	
	Science 287 (5461), 2185-2195 (2000)	
TITLE	20196006	
MEDLINE	2 (bases 1 to 280887)	
REFERENCE	Adams,M.D., Celniker,S.E., Gibbs,R.A., Rubin,G.M. and Venter,C.J.	
AUTHORS	Direct Submission	
TITLE	Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive.	
JOURNAL	Rockville, MD, USA	
COMMENT	On Oct 9, 2000 this sequence version replaced gi:7294043.	
FEATURES	Location/Qualifiers	
Source	1. 280887	
	/organism="Drosophila melanogaster"	
	/db_xref="taxon:7227"	
	/chromosome="3L"	
	join(81..401,546..679,736..1039,1142..2073)	
	/gene="CG4159"	
	/product="Ct13760"	
	/db_xref="FLYBASE:FBan0004159"	
	/db_xref="FLYBASE:FBgn0035642"	
	<81..>2073	
	/gene="CG4159"	
	/map="73A7-73A9"	
	/db_xref="FLYBASE:FBan0004159"	
	/db_xref="FLYBASE:FBgn0035642"	
	join(357..401,546..679,736..1039,1142..1981)	
	/gene="CG4159"	
	/note="CG4159 gene product"	
mrna		
	complement(join(<6102..6251,6309..6712,6961..>7000))	
	/gene="tra"	
	/note="Nucleotide sequence of the Celera sequence differs from the published sequence for this transcript."	
	/product="Ct137211"	
	/db_xref="FLYBASE:FBan0016724"	
	/db_xref="FLYBASE:FBgn0003741"	
	complement(<6102..>7000)	
	/gene="tra"	
	/note="CG16724"	
	/map="73A7-73A9"	
	/db_xref="FLYBASE:FBan0016724"	
	/db_xref="FLYBASE:FBgn0003741"	
CDS		
	complement(join(<6102..6251,6309..6712,6961..>7000))	
	/gene="tra"	
	/note="CG16724"	
	/map="73A7-73A9"	
	/db_xref="FLYBASE:FBan0016724"	
	/db_xref="FLYBASE:FBgn0003741"	
	complement(join(<6102..6251,6309..6712,6961..>7000))	
	/gene="tra"	
	/note="CG16724"	
	/map="73A7-73A9"	
	/db_xref="FLYBASE:FBan0016724"	
	/db_xref="FLYBASE:FBgn0003741"	
	complement(join(<6102..6251,6309..6712,6961..>7000))	
	/gene="tra"	
	/note="CG16724"	
	/map="73A7-73A9"	
	/db_xref="FLYBASE:FBan0016724"	
	/db_xref="FLYBASE:FBgn0003741"	
	complement(join(<6102..6251,6309..6712,6961..>7000))	
	/gene="tra"	
	/note="CG16724"	
	/map="73A7-73A9"	
	/db_xref="FLYBASE:FBan0016724"	
	/db_xref="FLYBASE:FBgn0003741"	
	complement(join(<6102..6251,6309..6712,6961..>7000))	
	/gene="tra"	
	/note="CG16724"	
	/map="73A7-73A9"	
	/db_xref="FLYBASE:FBan0016724"	
	/db_xref="FLYBASE:FBgn0003741"	
	complement(join(<6102..6251,6309..6712,6961..>7000))	
	/gene="tra"	
	/note="CG16724"	
	/map="73A7-73A9"	
	/db_xref="FLYBASE:FBan0016724"	
	/db_xref="FLYBASE:FBgn0003741"	
	complement(join(<6102..6251,6309..6712,6961..>7000))	
	/gene="tra"	
	/note="CG16724"	
	/map="73A7-73A9"	
	/db_xref="FLYBASE:FBan0016724"	
	/db_xref="FLYBASE:FBgn0003741"	
	complement(join(<6102..6251,6309..6712,6961..>7000))	
	/gene="tra"	
	/note="CG16724"	
	/map="73A7-73A9"	
	/db_xref="FLYBASE:FBan0016724"	
	/db_xref="FLYBASE:FBgn0003741"	
	complement(join(<6102..6251,6309..6712,6961..>7000))	
	/gene="tra"	
	/note="CG16724"	
	/map="73A7-73A9"	
	/db_xref="FLYBASE:FBan0016724	





JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNALREFERENCE  
AUTHORS

Unpublished  
2 (bases 1 to 181437)  
Worley, K.C.  
Submitted (22-SEP-1999) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 181437)

Alstbrooks, S.L., Amaralunga, H.C., Are, J.R., Banks, T., Barbarta, J.,  
Benton, J., Bimaye, M., Brown, E., Brown, M., Bryant, N.P., Bunay, C.,  
Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P., Bunay, C.,  
Burck, P., Burkett, C., Burrell, K.L., Byrd, N.C., Caron, T.F.,  
Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R.,  
Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C.,  
Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C.,  
Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O.,  
Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Drepper, H.,  
Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C.,  
Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J.,  
Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T.,  
Garza, N., Gill, R., Gorrell, J.H., Guevara, M., Gunaratne, P., Hale, S.,  
Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A.,  
Hernandez, J., Hernandez, O., Hodgson, A., Hogues, M., Holloway, C.,  
Hollins, B., Homs, F., Howard, S., Huber, J., Hulyk, S., Hume, J.,  
Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S.,  
Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J.,  
Kovar, C., Kralovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C.,  
Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W.,  
Loulsegad, H., Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R.,  
Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A.,  
Martinez, E., Massey, E., Maxhiney, E., McLeod, M.P., Meador, M.,  
Mel, G., Metzger, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K.,  
Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N.,  
Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nkwenkwo, S.,  
Ogih, M., Okuwonu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B.,  
Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L.,  
Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojibokan, I., Rolfe, M.,  
Ruiz, S., Saverly, G., Scherer, S., Scott, G., Shen, H., Shoshitari, N.,  
Sisson, I., Sodergren, E., Sonalke, T., Sparks, A., Stanley, H.,  
Stone, H., Sutton, A., Svatek, A., Taber, P., Tamerisa, A., Tamerisa, K.,  
Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N.,  
Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R.,  
Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C.,  
Washington, S., Williams, G., Williamson, A., Wlezyk, R., Wooden, S.,  
Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,  
Weinstock, G. and Gibbs, R.

TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
Direct Submission  
Submitted (10-NOV-2001) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
4 (bases 1 to 181437)

Alstbrooks, S.L., Amaralunga, H.C., Are, J.R., Banks, T., Barbarta, J.,  
Benton, J., Bimaye, M., Brown, E., Brown, M., Bryant, N.P., Bunay, C.,  
Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P., Bunay, C.,  
Burck, P., Burkett, C., Burrell, K.L., Byrd, N.C., Caron, T.F.,  
Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R.,  
Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C.,  
Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C.,  
Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O.,  
Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Drepper, H.,  
Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C.,  
Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J.,  
Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T.,  
Garza, N., Gill, R., Gorrell, J.H., Guevara, M., Gunaratne, P., Hale, S.,  
Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A.,  
Hernandez, J., Hernandez, O., Hodgson, A., Hogues, M., Holloway, C.,  
Hollins, B., Homs, F., Howard, S., Huber, J., Hulyk, S., Hume, J.,  
Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S.,  
Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J.,  
Kovar, C., Kralovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C.,  
Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W.,

JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNALREFERENCE  
AUTHORS

Loulsegad, H., Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R.,  
Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A.,  
Martinez, E., Massey, E., Maxhiney, E., McLeod, M.P., Meador, M.,  
Mel, G., Metzger, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K.,  
Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N.,  
Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nkwenkwo, S.,  
Ogih, M., Okuwonu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B.,  
Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L.,  
Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojibokan, I., Rolfe, M.,  
Ruiz, S., Saverly, G., Scherer, S., Scott, G., Shen, H., Shoshitari, N.,  
Sisson, I., Sodergren, E., Sonalke, T., Sparks, A., Stanley, H.,  
Stone, H., Sutton, A., Svatek, A., Taber, P., Tamerisa, A., Tamerisa, K.,  
Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N.,  
Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R.,  
Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C.,  
Washington, S., Williams, G., Williamson, A., Wlezyk, R., Wooden, S.,  
Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,  
Weinstock, G. and Gibbs, R.

JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

## COMMENT

Direct Submission  
Submitted (03-JAN-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
On Nov 10, 2001 this sequence version replaced gi:8101327.  
INFORMATION: <http://www.ngsc.bcm.tmc.edu/> or email  
gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the  
entire insert of this clone. Overlapping regions of clones are only  
sequenced and submitted once, so the sequence for the remainder of





*	14633	18437	contig of 3805 bp in length
*	14638	18537	gap of unknown length
*	18538	22123	contig of 3586 bp in length
*	22124	22223	gap of unknown length
*	22224	25731	contig of 3508 bp in length
*	25732	25831	gap of unknown length
*	25832	28890	contig of 3059 bp in length
*	28891	28990	gap of unknown length
*	28991	31876	contig of 2886 bp in length
*	31877	31977	gap of unknown length
*	31977	34726	contig of 2750 bp in length
*	34727	34826	gap of unknown length
*	34827	37144	contig of 2318 bp in length
*	37145	37244	gap of unknown length
*	37245	40483	contig of 3139 bp in length
*	40384	40483	gap of unknown length
*	40484	43322	contig of 2839 bp in length
*	43323	43322	gap of unknown length
*	43323	45399	contig of 1977 bp in length
*	43423	45499	gap of unknown length
*	45400	47906	contig of 2407 bp in length
*	45500	48006	gap of unknown length
*	47907	50859	contig of 2853 bp in length
*	48007	50860	gap of unknown length
*	50860	50859	gap of unknown length
*	50960	54178	contig of 3119 bp in length
*	54079	54178	gap of unknown length
*	54179	56138	contig of 1960 bp in length
*	56139	56238	gap of unknown length
*	56239	59204	contig of 2966 bp in length
*	59205	59305	gap of unknown length
*	59305	62250	contig of 2946 bp in length
*	62251	62350	gap of unknown length
*	62351	65162	contig of 2812 bp in length
*	65163	65262	gap of unknown length
*	65263	67095	contig of 1833 bp in length
*	67096	67195	gap of unknown length
*	67196	69359	contig of 2164 bp in length
*	69360	69459	gap of unknown length
*	72018	72018	contig of 2559 bp in length
*	72019	72118	gap of unknown length
*	72119	74575	contig of 2457 bp in length
*	74576	74675	gap of unknown length
*	74676	76752	contig of 2077 bp in length
*	76753	76852	gap of unknown length
*	76853	78542	contig of 1630 bp in length
*	78543	78642	gap of unknown length
*	78643	79814	contig of 1172 bp in length
*	79815	79914	gap of unknown length
*	79915	81212	contig of 1288 bp in length
*	81213	81312	gap of unknown length
*	81313	84330	contig of 3008 bp in length
*	84321	84420	gap of unknown length
*	84421	86815	contig of 2335 bp in length
*	86816	86915	gap of unknown length
*	86916	90121	contig of 3206 bp in length
*	90122	90221	gap of unknown length
*	90222	92334	contig of 2113 bp in length
*	92335	92434	gap of unknown length
*	92435	94033	contig of 1559 bp in length
*	94034	94133	gap of unknown length
*	94134	96100	contig of 1967 bp in length
*	96101	96200	gap of unknown length
*	96201	97631	contig of 1431 bp in length
*	97632	97731	gap of unknown length
*	97732	99389	contig of 1658 bp in length
*	99390	99489	gap of unknown length
*	99490	101608	contig of 2119 bp in length
*	101609	101708	gap of unknown length
*	101709	103260	contig of 1552 bp in length
*	103261	103360	gap of unknown length
*	103361	105942	contig of 2582 bp in length
*	105943	107862	gap of unknown length
*	106043	107786	contig of 1744 bp in length

[illegible]





Tue Apr 30 14:18:46 2002

```

; TOPOLOGY: linear
US-09-225-170-13

Query Match
Best Local Similarity 100.0%; Pred. No. 2;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 362 cattataaaaaaa 382
|||||
DB 4859 CATTAAAAA 4839

RESULT 26
US-08-441-139-12/c
; Sequence 12, Application US/08441139
; Patent No. 5773245
; GENERAL INFORMATION:
; APPLICANT: Wittup, Dr. Karl D.
; APPLICANT: Robinson, Anne S.
; TITLE OF INVENTION: METHODS FOR INCREASING SECRETION OF
; TITLE OF INVENTION: RECOMBINANTLY EXPRESSED PROTEINS
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCULLY, SCOTT, MURPHY & PRESSER
; STREET: 400 Garden City Plaza
; CITY: Garden City
; STATE: NY
; COUNTRY: USA
; ZIP: 11530
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/441,139
; FILING DATE: 15-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/089,997
; FILING DATE: 06-JUL-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Digiglio, Frank S.
; REGISTRATION NUMBER: 31,346
; REFERENCE/DOCKET NUMBER: 8646
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 516-742-4343
; TELEFAX: 516-742-4366
; TELEX: 230 501 SANS UR
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5470 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: exon
; LOCATION: 593..715
; FEATURE:
; NAME/KEY: exon
; LOCATION: 806..1036
; FEATURE:
; NAME/KEY: exon
; LOCATION: 1402..1539
; FEATURE:
; NAME/KEY: exon
; LOCATION: 2175..2289
; FEATURE:
; NAME/KEY: exon
; LOCATION: 2378..2764
; FEATURE:

; NAME/KEY: exon
; LOCATION: 2878..3115
; FEATURE:
; NAME/KEY: exon
; LOCATION: 3400..3568
; FEATURE:
; NAME/KEY: exon
; LOCATION: 4535..5095
; FEATURE:
US-08-441-139-12

Query Match
Best Local Similarity 100.0%; Pred. No. 2;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 361 ccattataaaaaaa 381
|||||
DB 4042 CCATTAAAAA 4022

RESULT 27
5196523-5/c
; Patent No. 5196523
; APPLICANT: LEE, AMY S.
; TITLE OF INVENTION: CONTROL OF GENE EXPRESSION BY GLUCOSE,
; CALCIUM AND TEMPERATURE
; NUMBER OF SEQUENCES: 28
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/354,988
; FILING DATE: 19-MAY-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 282,880
; FILING DATE: 05-DEC-1988
; APPLICATION NUMBER: 690,951
; FILING DATE: 01-JAN-1985
; SEQ ID NO: 5
; LENGTH: 5470
5196523-5

Query Match
Best Local Similarity 100.0%; Pred. No. 2;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 361 ccattataaaaaaa 381
|||||
DB 4042 CCATTAAAAA 4022

RESULT 28
US-08-070-501-2
; Sequence 2, Application US/08070501
; Patent No. 5671995
; GENERAL INFORMATION:
; APPLICANT: IIDA, Toshio
; APPLICANT: KAMINUMA, Toshihiko
; APPLICANT: FUSE, Yuka
; APPLICANT: TAJIMA, Masahiro
; APPLICANT: YAHAGI, Mitsuo
; APPLICANT: OKAMOTO, Hiroshi
; APPLICANT: KISHIMOTO, Jiro
; APPLICANT: IFUKU, Ohji
; APPLICANT: KATO, Ichiro
; TITLE OF INVENTION: ENZYME PARTICIPATING IN C-TERMINAL
; TITLE OF INVENTION: AMIDATION, AND METHOD OF PREPARING SAME AND USE THEREOF
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wegner, Cantor, Mueller & Player, P.C.
; STREET: 1233 20th Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20036-8218

```



```

;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/070,301
; FILING DATE: 24-MAY-1991
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 1-209687
; FILING DATE: 15-AUG-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 1-181933
; FILING DATE: 31-OCT-1985
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 2-76331
; FILING DATE: 26-MAR-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 2-106412
; FILING DATE: 24-APR-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 2-205475
; FILING DATE: 02-AUG-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Player, William E.
; REGISTRATION NUMBER: 31,409
; REFERENCE/DOCKET NUMBER: P-450-22830
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 887-040
; TELEFAX: (202) 835-0605
; TELEX: 440706
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6636 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: Horse
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 11..3070
; US-08-070-301-2

Query Match 5.5%; Score 21; DB 2; Length 6636;
Best Local Similarity 100.0%; Pred. No. 1.9;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 362 cattataaaaaaaaaaaaaa 382
Db 5186 CATTAAAAAaaaaaaaaaaaa 5206

RESULT 29
US-09-318-448-11/c
; Sequence 11, Application US/09318448
; Patent No. 6210950
; GENERAL INFORMATION:
; APPLICANT: Johnson, William G.
; APPLICANT: Stenroos, Edward S.
; TITLE OF INVENTION: METHODS FOR DIAGNOSING, PREVENTING, AND TREATING
; TITLE OF INVENTION: DEVELOPMENTAL DISORDERS
; FILE REFERENCE: 601-1-057
; CURRENT APPLICATION NUMBER: US/09/318,448
; CURRENT FILING DATE: 1999-05-25
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
; LENGTH: 18596
; TYPE: DNA

;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/070,301
; FILING DATE: 24-MAY-1991
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 1-209687
; FILING DATE: 15-AUG-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 1-181933
; FILING DATE: 31-OCT-1985
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 2-76331
; FILING DATE: 26-MAR-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 2-106412
; FILING DATE: 24-APR-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 2-205475
; FILING DATE: 02-AUG-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Player, William E.
; REGISTRATION NUMBER: 31,409
; REFERENCE/DOCKET NUMBER: P-450-22830
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 887-040
; TELEFAX: (202) 835-0605
; TELEX: 440706
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6636 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: Horse
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 11..3070
; US-08-070-301-2

Query Match 5.5%; Score 21; DB 2; Length 6636;
Best Local Similarity 100.0%; Pred. No. 1.9;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 362 cattataaaaaaaaaaaaaa 382
Db 5186 CATTAAAAAaaaaaaaaaaaa 5206

RESULT 29
US-09-318-448-11/c
; Sequence 11, Application US/09318448
; Patent No. 6210950
; GENERAL INFORMATION:
; APPLICANT: Johnson, William G.
; APPLICANT: Stenroos, Edward S.
; TITLE OF INVENTION: METHODS FOR DIAGNOSING, PREVENTING, AND TREATING
; TITLE OF INVENTION: DEVELOPMENTAL DISORDERS
; FILE REFERENCE: 601-1-057
; CURRENT APPLICATION NUMBER: US/09/318,448
; CURRENT FILING DATE: 1999-05-25
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
; LENGTH: 18596
; TYPE: DNA

;
; ORGANISM: Homo sapiens
; US-09-318-448-11

Query Match 5.5%; Score 21; DB 4; Length 18596;
Best Local Similarity 100.0%; Pred. No. 1.8;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 362 cattataaaaaaaaaaaaaa 382
Db 10785 CATTAAAAAaaaaaaaaaaaa 10765

RESULT 30
US-08-370-975B-6/c
; Sequence 6, Application US/08370975B
; Patent No. 5622851
; GENERAL INFORMATION:
; APPLICANT: Maley, Frank
; APPLICANT: Maley, Gladys F.
; APPLICANT: Weiner, Karen X.B.
; TITLE OF INVENTION: Human Deoxycytidylate Deaminase Gene
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle
; STREET: Clinton Square, P.O. Box 1051
; CITY: Rochester
; STATE: New York
; COUNTRY: USA
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/370,975B
; FILING DATE: 10-JAN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Timian, Susan J.
; REGISTRATION NUMBER: 34,103
; REFERENCE/DOCKET NUMBER: 20894/80
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716)263-1636
; TELEFAX: (716)263-1600
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20303 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT: 4q35
; US-08-370-975B-6

Query Match 5.5%; Score 21; DB 1; Length 20303;
Best Local Similarity 100.0%; Pred. No. 1.8;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 362 cattataaaaaaaaaaaaaa 382
Db 13143 CATTAAAAAaaaaaaaaaaaa 13123

RESULT 31
US-08-370-975B-1/c
; Sequence 1, Application US/08370975B
; Patent No. 5622851
; GENERAL INFORMATION:
; APPLICANT: Maley, Frank
```

APPLICANT: Maley, Gladys F.  
APPLICANT: Weiner, Karen X.B.  
TITLE OF INVENTION: Human Deoxycytidylate Deaminase Gene  
NUMBER OF SEQUENCES: 14  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Nixon, Hartgrave, Devans & Doyle  
STREET: Clinton Square, P.O. Box 1051  
CITY: Rochester  
STATE: New York  
COUNTRY: USA  
ZIP: 14603  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/370,975B  
FILING DATE: 10-JAN-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Timian, Susan J.  
REGISTRATION NUMBER: 34,103  
REFERENCE/DOCKET NUMBER: 20894/80  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (716)263-1636  
TELEFAX: (716)263-1600  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 26764 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
POSITION IN GENOME:  
CHROMOSOME/SEGMENT: 4q35  
US-08-370-975B-1

Query Match 5.5%; Score 21; DB 1; Length 26764;  
Best Local Similarity 100.0%; Pred. No. 1.7;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 362 cattataaaaaaaaaaaaaa 382  
Db 15106 cattataaaaaaaaaaaaaa 15086

RESULT 32  
US-08-113-646A-6  
Sequence 6, Application US/08113646A  
Patent No. 5578468  
GENERAL INFORMATION:  
APPLICANT: PICKUP, David J.  
APPLICANT: PATEL, Dhaval Kumar  
APPLICANT: ANTICZAK, James B.  
TITLE OF INVENTION: SITE-SPECIFIC RNA CLEAVAGE  
NUMBER OF SEQUENCES: 44  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: NIXON & VANDERHIVE P.C.  
STREET: 1100 NORTH GLEBE ROAD  
CITY: ARLINGTON  
STATE: VIRGINIA  
COUNTRY: U.S.A.  
ZIP: 22201-4714  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/113,646A  
FILING DATE: 31-AUG-1993

CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/084,406  
FILING DATE: 10-AUG-1987  
ATTORNEY/AGENT INFORMATION:  
NAME: WILSON, MARY J.  
REGISTRATION NUMBER: 32,955  
REFERENCE/DOCKET NUMBER: 1579-20  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 816-4100  
TELEFAX: (703) 816-4100  
TELEX: 200757 MIXX UR  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 36 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: RNA (genomic)  
US-08-113-646A-6

Query Match 5.2%; Score 20; DB 1; Length 36;  
Best Local Similarity 90.0%; Pred. No. 7.7;  
Matches 18; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 363 attataaaaaaaaaaaaaa 382  
Db 16 AUUAAAAAAAAAAAAAAAAA 35

RESULT 33  
US-08-799-464A-35  
Sequence 35, Application US/08799464A  
Patent No. 5958601  
GENERAL INFORMATION:  
APPLICANT: Murtaugh, Michael P. et al.  
TITLE OF INVENTION: VR-2332 VIRAL NUCLEOTIDE SEQUENCE AND METHODS OF USE  
NUMBER OF SEQUENCES: 35  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: John M. Collins  
STREET: 2405 Grand Blvd., Suite 400  
CITY: Kansas City  
STATE: Missouri  
COUNTRY: USA  
ZIP: 64108  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/795,464A  
FILING DATE:  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/287,941  
FILING DATE: August 5, 1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Collins, John M.  
REGISTRATION NUMBER: 26,262  
REFERENCE/DOCKET NUMBER: 22507  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (816) 474-9050  
TELEFAX: (816) 474-9057  
INFORMATION FOR SEQ ID NO: 35:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 135 bases  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA

```

: HYPOTHETICAL: NO
: US-08-799-464A-35

Query Match          5.2%; Score 20; DB 2; Length 135;
Best Local Similarity 100.0%; Pred. No. 6.9;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 363 attaaaaa382
    |||||
Db 113 ATTAAAAA382

RESULT 34
US-08-799-464A-34
: Sequence 34, Application US/08799464A
: Patent No. 5998601
: GENERAL INFORMATION:
: APPLICANT: Mutaugh, Michael P. et al.
: TITLE OF INVENTION: VR-2332 VIRAL NUCLEOTIDE SEQUENCE AND
: METHODS OF USE
: NUMBER OF SEQUENCES: 35
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: John M. Collins
: STREET: 2405 Grand Blvd., Suite 400
: CITY: Kansas City
: STATE: Missouri
: COUNTRY: USA
: ZIP: 64108
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/799,464A
: FILING DATE:
: CLASSIFICATION: 536
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/287,941
: FILING DATE: August 5, 1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Collins, John M.
: REGISTRATION NUMBER: 26,262
: REFERENCE/DOCKET NUMBER: 22907
: TELEPHONE: (816) 474-9050
: TELEFAX: (816) 474-9057
: INFORMATION FOR SEQ ID NO: 34:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 171 bases
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA
: HYPOTHETICAL: NO
: US-08-799-464A-34

Query Match          5.2%; Score 20; DB 2; Length 171;
Best Local Similarity 100.0%; Pred. No. 6.8;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 363 attaaaaa382
    |||||
Db 149 ATTAAAAA168

RESULT 35
US-08-721-488-9
: Sequence 9, Application US/08721488
: Patent No. 5965388
: GENERAL INFORMATION:
: APPLICANT: Jacobs, Kenneth
: APPLICANT: McCoy, John
: APPLICANT: LaVallie, Edward
: APPLICANT: Racie, Lisa
: APPLICANT: Merberg, David
: APPLICANT: Treacy, Maurice
: APPLICANT: Spaulding, Vikki
: APPLICANT: Bowman, Michael
: TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
: NUMBER OF SEQUENCES: 16
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Genetics Institute, Inc.
: STREET: 87 CambridgePark Drive
: CITY: Cambridge
: STATE: Massachusetts
: COUNTRY: U.S.A.
: ZIP: 02140
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/721,488
: FILING DATE:
: CLASSIFICATION: 424
: ATTORNEY/AGENT INFORMATION:
: NAME: Brown, Scott A.
: REGISTRATION NUMBER: 32,724
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (617) 498-8224
: TELEFAX: (617) 876-5651
: INFORMATION FOR SEQ ID NO: 9:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 358 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: US-08-721-488-9

Query Match          5.2%; Score 20; DB 2; Length 358;
Best Local Similarity 100.0%; Pred. No. 6.4;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 363 attaaaaa382
    |||||
Db 337 ATTAAAAA356

RESULT 36
US-08-117-080-13
: Sequence 13, Application US/08117080
: Patent No. 5482928
: GENERAL INFORMATION:
: APPLICANT: DE BOLLE, MIGUEL
: APPLICANT: BROEKAERT, WILLEM F
: APPLICANT: CAMMUE, BRUNO PA
: APPLICANT: VANDERLEYDEN, JOZEF
: APPLICANT: REES, SARAH B
: TITLE OF INVENTION: BIOCIDAL PROTEINS
: NUMBER OF SEQUENCES: 13
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: CUSHMAN DARBY & CUSHMAN
: STREET: 1100 NEW YORK AVENUE, NW; NINTH FLOOR, EAST
: CITY: WASHINGTON
: STATE: D.C.
: COUNTRY: USA
: ZIP: 20005-3918
: COMPUTER READABLE FORM:
```

```

: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/117,080
: FILING DATE:
: CLASSIFICATION: 530
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: PCT/GB92/00423
: FILING DATE: 10-MAR-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: GB 9105052.6
: FILING DATE: 11-MAR-1991
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: GB 9105684.6
: FILING DATE: 19-MAR-1991
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (202) 861 3000
: TELEX: 6714627 CUSH
: INFORMATION FOR SEQ ID NO: 13:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 433 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: ORIGINAL SOURCE:
: ORGANISM: FIGURE 7 BASE SEQUENCE MJ-AMP2
: US-08-117-080-13

```

```

Query Match          5.2%; Score 20; DB 1; Length 433;
Best Local Similarity 100.0%; Pred. No. 6.3;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 363 attaaaaaiaaiaaiaaiaaiaa 382
DB 413 ATTAAAAAIAAIAAIAAIAAIAA 432

```

```

RESULT 37
US-08-471-329-13
: Sequence 13, Application US/08471329
: Patent No. 5685048
: GENERAL INFORMATION:
: APPLICANT: DE BOLLE, MIGUEL
: APPLICANT: BROEKAERT, WILLEM F
: APPLICANT: CAMMUE, BRUNO PA
: APPLICANT: VANDERLEYDEN, JOZEF
: APPLICANT: REES, SARAH B
: TITLE OF INVENTION: BIOCIDAL PROTEINS
: NUMBER OF SEQUENCES: 13
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: CUSHMAN DABY & CUSHMAN
: STREET: 1100 NEW YORK AVENUE, NW; NINTH FLOOR, EAST
: STREET: TOWER
: CITY: WASHINGTON
: STATE: D.C.
: COUNTRY: USA
: ZIP: 20005-3918
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/471,329
: FILING DATE: 02-JUN-1995
: CLASSIFICATION: 800
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/117,080

```

```

: FILING DATE: 20-DEC-1993
: APPLICATION NUMBER: PCT/GB92/00423
: FILING DATE: 10-MAR-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: GB 9105052.6
: FILING DATE: 11-MAR-1991
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: GB 9105684.6
: FILING DATE: 19-MAR-1991
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (202) 861 3000
: TELEX: 6714627 CUSH
: INFORMATION FOR SEQ ID NO: 13:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 433 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: ORIGINAL SOURCE:
: ORGANISM: FIGURE 7 BASE SEQUENCE MJ-AMP2
: US-08-471-329-13

```

```

Query Match          5.2%; Score 20; DB 1; Length 433;
Best Local Similarity 100.0%; Pred. No. 6.3;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 363 attaaaaaiaaiaaiaaiaaiaa 382
DB 413 ATTAAAAAIAAIAAIAAIAAIAA 432

```

```

RESULT 38
US-08-915-142-13
: Sequence 13, Application US/08915142
: Patent No. 5942663
: GENERAL INFORMATION:
: APPLICANT: DE BOLLE, MIGUEL
: APPLICANT: BROEKAERT, WILLEM F
: APPLICANT: CAMMUE, BRUNO PA
: APPLICANT: VANDERLEYDEN, JOZEF
: APPLICANT: REES, SARAH B
: TITLE OF INVENTION: BIOCIDAL PROTEINS
: NUMBER OF SEQUENCES: 13
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: CUSHMAN DABY & CUSHMAN
: STREET: 1100 NEW YORK AVENUE, NW; NINTH FLOOR, EAST
: STREET: TOWER
: CITY: WASHINGTON
: STATE: D.C.
: COUNTRY: USA
: ZIP: 20005-3918
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/915,142
: FILING DATE: 20-AUG-1997
: CLASSIFICATION: 800
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/471,329
: FILING DATE: 02-JUN-1995
: APPLICATION NUMBER: US 08/117,080
: FILING DATE: 20-DEC-1993
: APPLICATION NUMBER: PCT/GB92/00423
: FILING DATE: 10-MAR-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: GB 9105052.6
: FILING DATE: 11-MAR-1991

```

```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9105684.6
; FILING DATE: 19-MAR-1991
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 861 3000
; TELEFAX: (202) 822-0944
; TELEX: 6714677 CUSH
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 433 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: FIGURE 7 BASE SEQUENCE MJ-AMP2
; US-08-915-142-13

Query Match 5.2%; Score 20; DB 2; Length 433;
Best Local Similarity 100.0%; Pred. No. 6.3;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 363 attataaataaataaataa 382
DB 413 ATTAAAAAATAAATAAATAA 432

RESULT 39
US-08-289-247B-2
; Sequence 2, Application US/08289247B
; Patent No. 5728579
; GENERAL INFORMATION:
; APPLICANT: Morrison, Briggs W.
; TITLE OF INVENTION: Detection and Treatment of Breast
; TITLE OF INVENTION: Cancer
; NUMBER OF SEQUENCES: 4
; ADDRESSEE: Clark & Elbing LLP
; STREET: 176 Federal Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2223
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 502 or 55SX
; OPERATING SYSTEM: MS-DOS (Version 5.0)
; SOFTWARE: WordPerfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/289,247B
; FILING DATE: August 11, 1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Bleker-Brady, Kristina
; REGISTRATION NUMBER: 39,109
; REFERENCE/DOCKET NUMBER: 00383/021001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 428-0200
; TELEFAX: (617) 428-7045
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 526
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-289-247B-2

Query Match 5.2%; Score 20; DB 2; Length 526;
Best Local Similarity 100.0%; Pred. No. 6.2;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 363 attataaataaataaataa 382
DB 504 ATTAAAAAATAAATAAATAA 523

RESULT 40
PCT-US95-09762-2
; Sequence 2, Application PC/TUS9509762
; GENERAL INFORMATION:
; APPLICANT: Morrison, Briggs W.
; TITLE OF INVENTION: Detection and Treatment of Breast
; TITLE OF INVENTION: Cancer
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 502 or 55SX
; OPERATING SYSTEM: MS-DOS (Version 5.0)
; SOFTWARE: WordPerfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/09762
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/289,247
; FILING DATE: August 11, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Clark, Paul T.
; REGISTRATION NUMBER: 30,162
; REFERENCE/DOCKET NUMBER: 00383/021001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 526
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; PCT-US95-09762-2

Query Match 5.2%; Score 20; DB 5; Length 526;
Best Local Similarity 100.0%; Pred. No. 6.2;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 363 attataaataaataaataa 382
DB 504 ATTAAAAAATAAATAAATAA 523

RESULT 41
US-09-040-984-70
; Sequence 70, Application US/09040984
; Patent No. 6210883
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS
; TITLE OF INVENTION: OF LUNG CANCER
; NUMBER OF SEQUENCES: 86
; CORRESPONDENCE ADDRESS:
; US-09-040-984-70
```



ADDRESSEE: Sheridan Ross P.C.





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 30, 2002, 10:50:11 ; Search time 728.83 Seconds

(without alignments)  
708.135 Million cell updates/sec

Title: us-09-248-178-60  
Perfect score: 602  
Sequence: 1 tgaagagccgcgcgtgag.....tcattcgaggagttcgaagac 602

Scoring table:

OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 930621 seqs, 428662619 residues

Word size : 10

Total number of hits satisfying chosen parameters: 284374

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

N.Geneseq\_ll01.\*  
1: /SIDSI/gcgdata/geneseq/geneseq/NA1980.DAT.\*  
2: /SIDSI/gcgdata/geneseq/geneseq/NA1981.DAT.\*  
3: /SIDSI/gcgdata/geneseq/geneseq/NA1982.DAT.\*  
4: /SIDSI/gcgdata/geneseq/geneseq/NA1983.DAT.\*  
5: /SIDSI/gcgdata/geneseq/geneseq/NA1984.DAT.\*  
6: /SIDSI/gcgdata/geneseq/geneseq/NA1985.DAT.\*  
7: /SIDSI/gcgdata/geneseq/geneseq/NA1986.DAT.\*  
8: /SIDSI/gcgdata/geneseq/geneseq/NA1987.DAT.\*  
9: /SIDSI/gcgdata/geneseq/geneseq/NA1988.DAT.\*  
10: /SIDSI/gcgdata/geneseq/geneseq/NA1989.DAT.\*  
11: /SIDSI/gcgdata/geneseq/geneseq/NA1990.DAT.\*  
12: /SIDSI/gcgdata/geneseq/geneseq/NA1991.DAT.\*  
13: /SIDSI/gcgdata/geneseq/geneseq/NA1992.DAT.\*  
14: /SIDSI/gcgdata/geneseq/geneseq/NA1993.DAT.\*  
15: /SIDSI/gcgdata/geneseq/geneseq/NA1994.DAT.\*  
16: /SIDSI/gcgdata/geneseq/geneseq/NA1995.DAT.\*  
17: /SIDSI/gcgdata/geneseq/geneseq/NA1996.DAT.\*  
18: /SIDSI/gcgdata/geneseq/geneseq/NA1997.DAT.\*  
19: /SIDSI/gcgdata/geneseq/geneseq/NA1998.DAT.\*  
20: /SIDSI/gcgdata/geneseq/geneseq/NA1999.DAT.\*  
21: /SIDSI/gcgdata/geneseq/geneseq/NA2000.DAT.\*  
22: /SIDSI/gcgdata/geneseq/geneseq/NA2001.DAT.\*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	602	100.0	602	AAH84206	DNA encoding human
2	602	100.0	602	AAH79435	CDNA sequence of h
3	456	75.7	1815	AAH78023	Human cancer assoc
c 4	18	3.0	459	AAH83407	P. chrysogenum ABC
c 5	18	3.0	459	AAH83408	Plant viral move
c 6	18	3.0	771	AAH79351	Human leukocyte ce
c 7	18	3.0	1113	AAH49553	Human prostate can
c 8	18	3.0	1557	AAH23458	C. thermoaminogene
c 9	18	3.0	1704	AAH45373	DNA encoding human
c 10	18	3.0	1770	AAH38502	Human EST-derived
c 11	18	3.0	1776	AAH98783	

12	18	3.0	1824	19	AAH49552	Human leukocyte ce
13	18	3.0	2655	22	AAH05226	Human secreted pro
c 14	18	3.0	2271	15	AAH02176	Human TLE-2 gene.
c 15	18	3.0	2271	15	AAH05334	Human TLE-2 gene.
c 16	18	3.0	7746	21	AAH07836	Human homologue of
c 17	17	2.8	458	20	AAH40583	Human secreted pro
c 18	17	2.8	385	21	AAH06962	Human secreted pro
c 19	17	2.8	413	21	AAH06983	Human secreted pro
c 20	17	2.8	477	21	AAH76172	Human ORFX ORF1727
c 21	17	2.8	672	21	AAH57874	Leek glossy8 (gl8)
c 22	17	2.8	872	21	AAH57875	Leek glossy8 (gl8)
c 23	17	2.8	951	22	AAH50961	Phosphatidylinositol
c 24	17	2.8	1020	21	AAH30601	Human G protein-co
c 25	17	2.8	1020	21	AAH30718	DNA encoding human
c 26	17	2.8	1194	22	AAH88323	S. spinosa DNA fra
c 27	17	2.8	1258	21	AAH93450	Human secreted pro
c 28	17	2.8	1901	15	AAH66178	Seven transmembran
c 29	17	2.8	1901	19	AAH18356	Human R12 seven tr
c 30	17	2.8	1901	21	AAH51725	Human 7TM receptor
c 31	17	2.8	1985	7	AAH60126	Sequence encoding
c 32	17	2.8	2010	22	AAH14321	Human CDNA sequenc
c 33	17	2.8	2453	18	AAH44092	Human G-protein th
c 34	17	2.8	2766	22	AAH66941	C glutamicum codin
c 35	17	2.8	2813	21	AAH75611	Human ORFX ORF1165
c 36	17	2.8	3092	22	AAH60348	Human polynucleoti
c 37	17	2.8	3092	22	AAH90050	Human bone marrow
c 38	17	2.8	3352	22	AAH58562	Human polynucleoti
c 39	17	2.8	3405	22	AAH89937	Human bone marrow
c 40	17	2.8	5713	21	AAH76360	Human ORFX ORF1515
c 41	17	2.8	5813	22	AAH57490	Human liver cell s
c 42	17	2.8	6411	19	AAH52730	Human hepatocyte n
c 43	17	2.8	16335	19	AAH52207	Streptococcus pneu
c 44	17	2.8	45624	22	AAH88315	S. spinosa DNA fra
c 45	17	2.8	50000	22	AAH88312	S. spinosa DNA fra

#### ALIGNMENTS

RESULT 1  
AAH84206  
ID AAX84206 standard; cDNA; 602 BP.  
XX  
XX AAX84206;  
AC  
XX 08-SEP-1999 (first entry)  
DI  
XX DNA encoding human breast tumour protein immunogenic fragment.  
DE  
XX Breast tumour protein; immunogenic fragment.  
KW Breast cancer development; therapy; ss.  
XX  
XX Homo sapiens.  
OS  
XX  
XX W05933869-A2.  
XX  
XX 08-JUL-1999.  
XX  
XX 22-DEC-1998; 98WO-US/7416.  
XX  
XX 17-JUL-1998; 98US-018627.  
PR 24-DEC-1997; 97US-0998253.  
PR 24-DEC-1997; 97US-0998255.  
PR 17-JUL-1998; 98US-0118554.  
XX  
XX (CORI-) CORTIXA CORP.  
XX  
XX Reed SG, Xu J;  
XX WPI; 1999-405486/34.  
XX New breast tumour protein genes used, in vaccines for immunotherapy,  
PT or for diagnosis of breast cancer

XX PS Claim 3; Page 59; 70pp; English.  
XX CC This sequence encodes a human breast tumour protein immunogenic fragment  
CC of the invention. The polypeptides or nucleic acids encoding them are  
CC useful in vaccines and pharmaceutical compositions for manufacture of  
CC medicaments for inhibiting the development of breast cancer in a patient.  
CC They can also be used to treat breast cancer. Antibodies against these  
CC polypeptides can be used to detect and monitor progression of breast  
CC cancer in patients. Primers and probes derived from the polynucleotides  
CC encoding the breast proteins are useful for detection of breast cancer.  
CC Peripheral blood cells from a patient incubated in the presence of at  
CC least one polypeptide, such that T cells proliferate, are useful in  
CC manufacture of a medicament for treating breast cancer in a patient.  
CC Antigen presenting cells incubated in the presence of at least one  
CC polypeptide are also useful for treating breast cancer.  
XX SQ Sequence 602 BP; 147 A; 155 C; 192 G; 108 T; 0 other;

Query Match 100.0%; Score 602; DB 20; Length 602;  
Best Local Similarity 100.0%; Pred. No. 9, 4e-289;  
Matches 602; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 tgaagagccgcggtgagctgctcccgatggagactgccaaccttcccaagctgcagc 60  
DB 1 tgaagagccgcggtgagctgctcccgatggagactgccaaccttcccaagctgcagc 60  
QY 61 ttgtgtgagaatagtcgccagcgggtccacttggcgggtcagtgaggagaagcacc 120  
DB 61 ttgtgtgagaatagtcgccagcgggtccacttggcgggtcagtgaggagaagcacc 120  
QY 121 ggtccatctctgtgagctacccagcactccaaagctcagatgacagagctggagt 180  
DB 121 ggtccatctctgtgagctacccagcactccaaagctcagatgacagagctggagt 180  
QY 181 ctctcagcggctggcagagatcccaagactgcacagagtgccggcggtctgtagg 240  
DB 181 ctctcagcggctggcagagatcccaagactgcacagagtgccggcggtctgtagg 240  
QY 241 agcccgaggaaggaggtgctatgaagcagctgagctgagagctgagagcacc 300  
DB 241 agcccgaggaaggaggtgctatgaagcagctgagctgagagctgagagcacc 300  
QY 301 gagatgtgtcccggtctgctacacccagcagctatcctggagatcgtggcaacatccgga 360  
DB 301 gagatgtgtcccggtctgctacacccagcagctatcctggagatcgtggcaacatccgga 360  
QY 361 agcagaaggagagatgatacccaagatcttctgatacgaaggagcttccagaggaaatca 420  
DB 361 agcagaaggagagatgatacccaagatcttctgatacgaaggagcttccagaggaaatca 420  
QY 421 actccatctctggaaagctggacggacgttttcgggtgactgagctgtgttcaagg 480  
DB 421 actccatctctggaaagctggacggacgttttcgggtgactgagctgtgttcaagg 480  
QY 481 atgccagaaggagcagctgtcttccgaagccctataagatctagctgtcttcacagaga 540  
DB 481 atgccagaaggagcagctgtcttccgaagccctataagatctagctgtcttcacagaga 540  
QY 541 actcagcagctctatccagaccatccagagacacagcagccaccatcatcgcgaggttcgag 600  
DB 541 actcagcagctctatccagaccatccagagacacagcagccaccatcatcgcgaggttcgag 600  
QY 601 ac 602  
DB 601 ac 602

RESULT 2  
AAC75435  
ID AAC75435 standard; cDNA; 602 BP.  
XX

AC AAC75435;  
XX 07-FEB-2001 (first entry)  
LI cDNA sequence of human breast tumour clone 10166a.  
DE Human; breast tumour antigen; cytostatic; immunotherapy;  
KW breast cancer; vaccine; ss.  
XX Homo sapiens.  
XX WC200061756-A2.  
XX 19-OCT-2000.  
XX 10-APR-2000; 2000WO-050766.  
XX 09-APR-1999; 5905-020950.  
XX 02-JUL-1999; 5905-0346327.  
XX (CORI-) CORIXA CORP.  
XX Reed SG, Xu J, Dillman DC;  
XX WPI: 2000-636556/61.  
XX A novel isolated polypeptide comprising an immunogenic portion of a  
XX breast cancer protein useful in the detection and treatment of breast  
XX cancer.  
XX Claim 4; Page 75-76; 55pp; English.  
XX The present sequence was isolated from a breast tumour cDNA library. It  
XX is provided in a specification relating to compounds for immunotherapy  
XX and diagnosis of breast cancer. Breast tumour antigens and the  
XX polynucleotides that encode them may be used in the production of a  
XX pharmaceutical composition to be used in the treatment of breast cancer.  
XX Proliferated T cells and incubated antigen presenting cells are also  
XX required. The polypeptides and polynucleotides may also be used to  
XX produce a vaccine.  
XX Sequence 602 BP; 147 A; 155 C; 192 G; 108 T; 0 other;  
SQ  
Query Match 100.0%; Score 602; DB 21; Length 602;  
Best Local Similarity 100.0%; Pred. No. 9, 4e-289;  
Matches 602; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 tgaagagccgcggtgagctgctcccgatggagactgccaaccttgcgaagctgcagc 60  
DB 1 tgaagagccgcggtgagctgctcccgatggagactgccaaccttgcgaagctgcagc 60  
QY 61 ttgtgtgagaatagtcgccagcgggtccacttggcgggtcagtgaggagaagcacc 120  
DB 61 ttgtgtgagaatagtcgccagcgggtccacttggcgggtcagtgaggagaagcacc 120  
QY 121 ggtccatctctgtgagctacccagcactccaaagctcagatgacagagctggagt 180  
DB 121 ggtccatctctgtgagctacccagcactccaaagctcagatgacagagctggagt 180  
QY 181 ctctcagcggctggcagagatcccaagactgcacagagtgccggcggtctgtagg 240  
DB 181 ctctcagcggctggcagagatcccaagactgcacagagtgccggcggtctgtagg 240  
QY 241 agcccgaggaaggaggtgctatgaagcagctgagctgagagctgagagcacc 300  
DB 241 agcccgaggaaggaggtgctatgaagcagctgagctgagagctgagagcacc 300  
QY 301 gagatgtgtcccggtctgctacacccagcagctatcctggagatcgtggcaacatccgga 360  
DB 301 gagatgtgtcccggtctgctacacccagcagctatcctggagatcgtggcaacatccgga 360  
QY 361 agcagaaggagagatgatacccaagatcttctgatacgaaggagcttccagaggaaatca 420

Db 361 agcagaagagagatcaccaagattgtctgatcagaaggagcttcagaagaaatca 420  
 Qy 421 actcctctcgtggaagctgaccgagcttctggtgactgatgagcttctgttcaagg 480  
 Db 421 actcctctcgtggaagctgaccgagcttctggtgactgatgagcttctgttcaagg 480  
 Qy 481 atgccaaagagagatgcttctcgaagcctataagctactagctgctcgcacaga 540  
 Db 481 atgccaaagagagatgcttctcgaagcctataagctactagctgctcgcacaga 540  
 Qy 541 actgcagcagctcatccagaccatcagagacacagcagcaccatcatcgaggagttcgag 600  
 Db 541 actgcagcagctcatccagaccatcagagacacagcagcaccatcatcgaggagttcgag 600  
 Qy 601 ac 602  
 Db 601 ac 602

RESULT 3  
 AAC78023  
 ID AAC78023 standard; cDNA; 1815 BP.  
 XX  
 AC  
 AC  
 XX  
 DT 08-FEB-2001 (first entry)  
 XX  
 DE Human cancer associated gene sequence SEQ ID NO:417.  
 XX  
 KW Human; cancer associated gene; cancer antigen; detection; cancer;  
 diagnosis; cytostatic; proliferative; vulnerable; immunomodulator;  
 antidiabetic; antisthmatic; antirheumatic; antiarthritic; antiviral;  
 antiinflammatory; antithyroid; antiallergic; antibacterial; cardiant;  
 dermatological; neuroprotective; thrombolytic; coagulant; neotropic;  
 vasotropic; antipsoriatic; antiangiogenic; gene therapy; inflammation;  
 immune disorder; haematopoietic cell disorder; autoimmune disorder;  
 allergic reaction; graft versus host disease; organ rejection;  
 haemostatic; thrombolytic; cardiovascular disorder; infection;  
 neurological disease; drug screening; ss.  
 KW Homo sapiens.  
 OS  
 XX  
 FN WO200055350-A1.  
 XX  
 PD 21-SEP-2000.  
 XX  
 PF 08-MAR-2000; 2000WO-US05882.  
 XX  
 PR 12-MAR-1999; 99US-0124270.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Rosen CA, Ruben SM;  
 XX  
 DR WPI: 2000-587533/55.  
 XX  
 DR P-PSDB: AAB43814.  
 XX  
 PT Novel isolated nucleic acids comprising sequences encoding peptides  
 useful for treating or diagnosing e.g. cancer -  
 XX  
 PS Claim 1; Page 958-959; 2352pp; English.  
 XX  
 CC AAC77607 to AAC78448 encode the human cancer associated proteins given  
 CC in AAB43398 to AAB44239. The proteins can have activities based on the  
 CC tissues and cells the genes are expressed in. Example of activities  
 CC include: cytostatic; proliferative; vulnerable; immunomodulator;  
 CC antidiabetic; antisthmatic; antirheumatic; antiarthritic;  
 CC antiinflammatory; antithyroid; antiallergic; antibacterial; antiviral;  
 CC dermatological; neuroprotective; cardiant; thrombolytic; coagulant;  
 CC neotropic; vasotropic; antipsoriatic and antiangiogenic. The  
 CC polynucleotides and polypeptides can be used for preventing, treating or  
 CC ameliorating medical conditions and diagnosing pathological conditions.

CC Polynucleotides, polypeptides, antibodies, agonists and antagonists from  
 CC the present invention may be used to treat immune disorders by activating  
 CC or inhibiting the proliferation, differentiation or mobilisation of  
 CC immune cells, to treat disorders of haematopoietic cells, autoimmune  
 CC disorders, allergic reactions, graft versus host disease and organ  
 CC rejection, modulate haemostatic or thrombolytic activity, modulate  
 CC inflammation, cancers, cardiovascular disorders, neurological disease an:  
 CC bacterial or viral infections. The peptides, nucleotides, antibodies,  
 CC agonists and antagonists may be also be used in drug screens. AAC78448 to  
 CC AAC78457 and AAB44240 represent sequences used in the exemplification of  
 CC the present invention.  
 XX  
 SQ Sequence 1815 BP; 372 A; 477 C; 613 G; 350 T; 3 other;

Query Match 75.7%; Score 456; DB 21; Length 1815;  
 Best Local Similarity 100.0%; Pred. No. 2.2e-216;  
 Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 147 ctccgaagctgcagagattcagagagctggaattctctcgacgctggcagagatccaa 206  
 Db 599 ctccgaagctgcagagattcagagagctggaattctctcgacgctggcagagatccaa 658  
 Qy 207 gaactgcaccagagtgctccggcggtgctggaagagcccgaggaaggaggtctat 266  
 Db 655 gaactgcaccagagtgctccggcggtgctggaagagcccgaggaaggaggtctat 718  
 Qy 267 aagcagctgagtgcagagctggagagactctgccagagatgtgtcccgctggcctacac 326  
 Db 719 aagcagctgagtgcagagctggagagactctgccagagatgtgtcccgctggcctacac 778  
 Qy 327 cagcgctctcgtgagctgctggcacaatccggaagcagaggaagagatccacaaatc 386  
 Db 779 cagcgctctcgtgagctgctggcacaatccggaagcagaggaagagatccacaaatc 838  
 Qy 387 ttgtctgatacgaagagcttcagaagaaatcaactccctctatctgggaagctggaccgg 446  
 Db 839 ttgtctgatacgaagagcttcagaagaaatcaactccctctatctgggaagctggaccgg 898  
 Qy 447 acgtttcggctgactgagcttgcattgattcaagagatcccaagagacgagctgttcgg 506  
 Db 899 acgtttcggctgactgagcttgcattgattcaagagatcccaagagacgagctgttcgg 958  
 Qy 507 aagcctataagttatctgactgctctgcagagaaactgcagcagctcattccagaccatc 566  
 Db 959 aagcctataagttatctgactgctctgcagagaaactgcagcagctcattccagaccatc 1018  
 Qy 567 gaggacacagggcaccatcagcgggaggttcagagac 602  
 Db 1019 gaggacacagggcaccatcagcgggaggttcagagac 1054

RESULT 4  
 AAF83407/c  
 ID AAF83407 standard; cDNA; 459 BP.  
 XX  
 AC  
 AC  
 XX  
 DT 09-JUL-2001 (first entry)  
 XX  
 DE P. chrysogenum ABC transporter dd062 partial cDNA.  
 XX  
 KW Beta-lactam; micro-organism; ABC transporter; ATP-binding cassette;  
 KW adenosine triphosphate-binding cassette; antibiotic; penicillin;  
 KW cephalosporin; dd062; ss.  
 OS Penicillium chrysogenum.  
 XX  
 PN WO200132904-A2.  
 XX  
 PD 10-MAY-2001.  
 XX  
 PF 03-NOV-2000; 2000WO-EP11465.

XX 03-NOV-1999; 99EP-0203684.  
 PR 03-NOV-1999; 99EP-0203685.  
 PR 03-NOV-1999; 99EP-0203687.  
 PR 03-NOV-1999; 99EP-0203688.  
 PR 03-NOV-1999; 99EP-0203689.  
 PR 03-NOV-1999; 99EP-0203690.  
 PR 03-NOV-1999; 99EP-0203691.  
 PR 03-NOV-1999; 99EP-0203692.  
 PR 03-NOV-1999; 99EP-0203693.  
 PR 03-NOV-1999; 99EP-0203694.  
 XX (STAM ) DSM NV.  
 PA  
 XX Van Den Berg MA, Bovenberg RAL, Driessen AJM, Konings WN;  
 PI Schuurs TA, Nieboer M, Westerlaken I;  
 XX WPI; 2001-291055/30.  
 DR  
 XX Enhancing secretion of beta-lactam compounds from a micro-organism by  
 PT enhancing adenosine triphosphate-binding cassette transporter activity,  
 PT useful for producing e.g. penicillin and cephalosporins -  
 XX  
 XX Claim 9; Page 116; 116pp; English.  
 XX The invention relates to a method for enhancing the secretion of beta-  
 CC lactam compounds from a micro-organism that comprises enhancing ABC  
 CC (adenosine triphosphate (ATP)-binding cassette) transporter activity of  
 CC the micro-organism. The method is used for enhancing the production and  
 CC secretions of beta-lactam antibiotics like penicillin and cephalosporins.  
 CC Manipulation of the ABC-transport protein system provides a means for  
 CC enhancing beta-lactam secretion. The present sequence represents a  
 CC partial cDNA sequence of the P. chrysogenum ABC transporter dd062.  
 XX  
 SQ Sequence 459 BP; 100 A; 133 C; 119 G; 107 T; 0 other;

Query Match 3.0%; Score 18; DB 22; Length 459;  
 Best Local Similarity 100.0%; Pred. No. 30;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 534 caccgagaactgcagccag 551  
 |||||  
 Db 160 CACGAGAACTGCAGCCAG 143

RESULT 5  
 AAF83408/c  
 ID AAF83408 standard; cDNA; 459 BP.  
 XX  
 AC AAF83408;  
 DT 09-JUL-2001 (first entry)  
 DE P. chrysogenum ABC transporter dd062 cDNA.  
 XX Beta-lactam; micro-organism; ABC transporter; ATP-binding cassette;  
 KW adenosine triphosphate-binding cassette; antibiotic; penicillin;  
 KW cephalosporin; dd062; ss.  
 XX Penicillium chrysogenum.  
 OS  
 XX Key Location/Qualifiers  
 FH CDS 4..457  
 FT /\*tag= a  
 FT /product= "dd062"  
 FT /note= "the start and stop codons are not indicated"  
 PN WO200132904-A2.  
 XX  
 PD 10-MAY-2001.  
 XX  
 PF 03-NOV-2000; 2000WO-EP11489.

XX 03-NOV-1999; 99EP-0203684.  
 PR 03-NOV-1999; 99EP-0203685.  
 PR 03-NOV-1999; 99EP-0203687.  
 PR 03-NOV-1999; 99EP-0203688.  
 PR 03-NOV-1999; 99EP-0203689.  
 PR 03-NOV-1999; 99EP-0203690.  
 PR 03-NOV-1999; 99EP-0203691.  
 PR 03-NOV-1999; 99EP-0203692.  
 PR 03-NOV-1999; 99EP-0203693.  
 PR 03-NOV-1999; 99EP-0203694.  
 XX (STAM ) DSM NV.  
 PA  
 XX Van Den Berg MA, Bovenberg RAL, Driessen AJM, Konings WN;  
 PI Schuurs TA, Nieboer M, Westerlaken I;  
 XX WPI; 2001-291055/30.  
 DR  
 XX Enhancing secretion of beta-lactam compounds from a micro-organism by  
 PT enhancing adenosine triphosphate-binding cassette transporter activity,  
 PT useful for producing e.g. penicillin and cephalosporins -  
 XX  
 XX Claim 9; Page 116-117; 116pp; English.  
 XX The invention relates to a method for enhancing the secretion of beta-  
 CC lactam compounds from a micro-organism that comprises enhancing ABC  
 CC (adenosine triphosphate (ATP)-binding cassette) transporter activity of  
 CC the micro-organism. The method is used for enhancing the production and  
 CC secretions of beta-lactam antibiotics like penicillin and cephalosporins.  
 CC Manipulation of the ABC-transport protein system provides a means for  
 CC enhancing beta-lactam secretion. The present sequence represents the  
 CC cDNA sequence of the P. chrysogenum ABC transporter dd062.  
 XX  
 SQ Sequence 459 BP; 100 A; 133 C; 119 G; 107 T; 0 other;

Query Match 3.0%; Score 18; DB 22; Length 459;  
 Best Local Similarity 100.0%; Pred. No. 30;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 534 caccgagaactgcagccag 551  
 |||||  
 Db 160 CACGAGAACTGCAGCCAG 143

RESULT 6  
 AAC79351  
 ID AAC79351 standard; cDNA; 771 BP.  
 XX  
 AC AAC79351;  
 DT 06-FEB-2001 (first entry)  
 DE Plant viral movement protein encoding cDNA SEQ ID 7.  
 XX Plant viral movement protein; transport; transgenic plant;  
 KW viral resistance; cosuppression; ss.  
 XX Hevea brasiliensis.  
 OS  
 XX WO2000060088-A2.  
 PN  
 PD 12-OCT-2000.  
 XX  
 PF 06-APR-2000; 2000WO-US05110.  
 XX  
 PR 07-APR-1999; 59US-0128092.  
 XX  
 PA (DUPC ) DU PONT DE NEMOURS & CO E I.  
 XX  
 PI Krebbers E, Weng 2, Cahoon RE;

XX  
DR WPI: 2000-638467/61.  
DR P-PSDB; AAB44497.  
XX  
XX  
PI Novel viral movement polypeptides and polynucleotides useful in field  
PI of plant molecular biology, for producing transgenic plants, to prepare  
PI antibodies and in immunological screening of cDNA expression libraries  
PI  
XX  
XX  
PS Claim 2: Page 37; 62pp; English.  
XX  
XX Polynucleotide sequences AAC79348-C79375 encode plant viral movement  
CC proteins AAB4494-B44520. Some plant viruses have been shown to be able  
CC to establish systemic infections via movement proteins that utilize  
CC existing plant pathways to traffic macromolecules to surrounding cells.  
CC proteins such as those of the invention are similar to viral movement  
CC proteins that function in the transport of nucleic acids from cell to  
CC cell. The plant viral movement proteins are useful for obtaining a  
CC nucleic acid fragment encoding a viral movement protein. Polynucleotides  
CC encoding the plant viral movement proteins are useful for positive  
CC selection of a transformed cell. The proteins are useful in the field of  
CC plant molecular biology, and in the preparation of antibodies against the  
CC proteins. The proteins are also useful for isolating cDNAs and genes  
CC encoding homologous proteins from the same or other plant species, and to  
CC create transgenic plants in which the protein is presented at higher or  
CC lower levels than normal or in cell types or developmental stages in  
CC which they are not normally found. The proteins and nucleotide sequences  
CC may be used to control cosuppression and engineer plant virus  
CC resistance.  
XX  
XX Sequence 771 BP; 216 A; 138 C; 176 G; 236 T; 5 other;  
SQ

Query Match 3.0%; Score 18; DB 21; Length 771;  
Best Local Similarity 100.0%; Pred. No. 30;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 194 ggcagagatccagaact 211  
IIIIIIIIIIIIIIIIII  
DE 76 ggcagagatccagaact 93

RESULT 7  
AAV49553  
ID AAV49553 standard; cDNA to mRNA; 1113 BP.  
XX  
XX  
AC AAV49553;  
XX  
XX 21-OCT-1998 (first entry)  
XX  
XX Human leukocyte cell clone HP00804 cDNA #2.  
DE  
XX  
XX Transmembrane domain; human; nutrition; cytokine; cell proliferation;  
KW differentiation; immune system; stimulator; suppressor; regulator;  
KW hematopoiesis; activin; inhibitor; chemotactic; chemokinetic; receptor;  
KW haemostatic; thrombolytic; ligand; anti-inflammatory; tumour; ds.  
XX  
XX Homo sapiens.  
OS  
XX  
XX WO9821328-A2.  
PN  
XX  
XX 22-MAY-1998.  
PD  
XX  
XX 07-NOV-1997; 97WO-JP04056.  
PF  
XX  
XX 13-NOV-1996; 96JP-0301429.  
PR  
XX  
XX (PROT-) PROTEGENE INC.  
PA (SAGA) SAGAMI CHEM RES CENTRE.  
PA  
XX Kato S, Kobayashi M, Sekine S, Yamaguchi T;  
PI WPI: 1998-297932/26.  
XX  
XX

DR P-PSDB; AAW64535.  
XX  
XX Human protein having transmembrane domain - useful for, e.g.  
PI research and nutrition  
PI  
XX Claim 3: Page 118-119; 205pp; English.  
XX  
XX AAV49550-V49599 are cDNA sequences which encode human proteins containing  
CC a transmembrane domain. These proteins can be used for, e.g. research  
CC and nutrition, and may have cytokine and cell  
CC proliferation/differentiation, immune stimulating/suppressing,  
CC haematopoiesis regulating, tissue growth, activin/inhibin,  
CC chemotactic/chemokinetic, haemostatic and thrombolytic, receptor/ligand,  
CC anti-inflammatory or tumour inhibition activity.  
XX  
SQ Sequence 1113 BP; 209 A; 403 C; 264 G; 237 T; 0 other;  
SQ

Query Match 3.0%; Score 18; DB 19; Length 1113;  
Best Local Similarity 100.0%; Pred. No. 30;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 330 cgcattccctggagatcctg 347  
IIIIIIIIIIIIIIIIIIII  
DB 916 cgcattccctggagatcctg 933

RESULT 8  
AAZ33458  
ID AAZ33458 standard; cDNA; 1557 BP.  
XX  
XX AAZ33458;  
AC  
XX  
XX 08-DEC-1999 (first entry)  
XX  
XX Human prostate cancer-associated EST 36.  
DE  
XX  
XX Expressed sequence tag; EST; prostate tumor; antitumor; treatment;  
KW gene therapy; tissue specificity human; ss.  
KW  
XX Homo sapiens.  
OS  
XX  
XX DE19611193-A1.  
PN  
XX  
XX 16-SEP-1999.  
PD  
XX  
XX 10-MAR-1998; 98DE-1011193.  
PF  
XX  
XX 10-MAR-1998; 98DE-1011193.  
PR  
XX  
XX (META-) METAGEN GES GENOMFORSCHUNG MBH.  
PA  
XX  
XX Specht T, Hinzmann B, Schmitt A, Pillarsky C, Dähl E, Rosenthal A;  
PI WPI: 1999-519628/44.  
XX  
XX P-PSDB; AAY48255.  
DR  
XX  
XX New nucleic acid expressed at high level in prostatic tumor tissue and  
PI encoded polypeptides, useful for treating cancer and screening for  
PI therapeutic agents  
PI  
XX  
XX Claim 3; 97-98; 166pp; German.  
PS  
XX  
XX This invention describes novel nucleic acid sequences (A) that are  
CC expressed at high level in prostatic tumor tissue and encode gene  
CC products or their fragments. The products of the invention have  
CC antitumor activity. Polypeptides (I) encoded by (A) are used: (i) for  
CC identifying agents for treatment of prostatic cancer and (ii) for  
CC therapy of prostate cancer, optionally where expressed by gene therapy  
CC methods. (A) is also used to isolate full-length genes (for gene therapy)  
CC and for recombinant production of (I), which can be used to raise  
CC specific antibodies. (A) are identified by assembly of ESTs (expressed  
CC sequence tags) before they are analyzed for expression pattern (tissue

RESULT	10
AAV38502	
ID	AAV38502 standard; DNA; 1770 BP.
XX	
AC	AAV38502;

12-OCT-2001 (first entry)

us-09-248-178-60.rng

Tue Apr 30 14:18:49 2002

```

XX DE Human EST-derived coding sequence SEQ ID NO: 640.
XX KW Human: sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;
XX KW tomato; monkey; dog; sea urchin; expressed sequence tag; EST;
XX KW diagnostics; forensic test; gene mapping; genetic disorder;
XX KW biodiversity; gene therapy; nutrition; ss.
XX OS Homo sapiens.
XX PN WO200134477-A2.
XX PD 02-AUG-2001.
XX PF 25-JAN-2001; 2001WO-US02687.
XX PR 25-JAN-2000; 2000US-0491404.
XX PR 17-JUL-2000; 2000US-0617746.
XX PR 03-AUG-2000; 2000US-0631451.
XX PR 15-SEP-2000; 2000US-0663870.
XX PA (HYSE-) HYSEQ INC.
XX PI Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;
XX PI Cao Y, Drmanac RA, Zhang J, Werhman T;
XX PI WPI: 2001-476164/51.
XX PR P-PSDB; AAM24124.
XX PT Isolated polypeptide for treatment of diseases, diagnostics, raising
XX PT antibodies and research use.
XX PS Claim 1; Page 610-611; 1275pp; English.
XX CC The present invention provides the protein and coding sequences of novel
XX CC proteins from a variety of organisms, including human, dog, cat, horse,
XX CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea
XX CC urchin and tomato. These were derived from expressed sequence tags (ESTs)
XX CC from the organism of interest. They can be used in diagnostics,
XX CC forensics, gene mapping, identification of mutations, to assess
XX CC biodiversity and for nutritional purposes. The present sequence is a cDNA
XX CC of the invention.
XX CC Sequence 1776 BP; 323 A; 635 C; 456 G; 362 T; 0 other;
XX SQ

Query Match 3.0%; Score 18; DB 22; Length 1776;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 330 cgcacctcggagatcgctg 347
DB 992 cgcacctcggagatcgctg 1009

RESULT 12
AAV49552
ID AAV49552 standard; cDNA to mRNA; 1824 BP.
XX AC AAV49552;
XX AC AAV49552;
XX DT 21-OCT-1998 (first entry)
XX XX Human leukocyte cell clone HP00804 cDNA #1.
XX DE Transmembrane domain; human; nutrition; cytokine; cell proliferation;
XX KW differentiation; immune system; stimulator; suppressor; regulator;
XX KW hematopoiesis; activin; inhibitor; chemotactic; chemokinetic; receptor;
XX KW haemostatic; thrombolytic; ligand; anti-inflammatory; tumour; ds.
XX XX Homo sapiens.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers

Query Match 3.0%; Score 18; DB 19; Length 1824;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 330 cgcacctcggagatcgctg 347
DB 1048 cgcacctcggagatcgctg 1065

RESULT 13
AA05226
ID AA05226 standard; cDNA; 2055 BP.
XX AC AA05226;
XX AC AA05226;
XX DT 18-JUL-2001 (first entry)
XX XX Human secreted protein-encoding gene 7 cDNA clone HCHP82, SEQ ID NO:17.
XX DE Human: secreted protein; proliferative disorder; cancer; tumour; asthma;
XX KW fetal abnormality; developmental abnormality; haematopoietic disorder;
XX KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
XX KW Parkinson's disease; cognitive disorder; schizophrenia; skin disorder;
XX KW psoriasis; sepsis; diabetes; atherosclerosis; cardiovascular disorder;
XX KW inflammation; neurological disorder; Alzheimer's disease; food additive;
XX KW angiogenic disorder; kidney disorder; gastrointestinal disorder; allergy;
XX KW pregnancy-related disorder; endocrine disorder; infection; wound healing;
XX KW cell culture; chemotaxis; vulnery; binding partner identification;
XX KW gene therapy; ss.
XX XX Homo sapiens.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX FH CDS 369..488
XX FT /tag= a
XX FT /product= "Human secreted protein"
XX FT sig_peptide 369..449

```

1



us-09-248-178-60.rng

Tue Apr 30 14:18:49 2002

```

DE Human TLE-2 gene.
XX TLE-2; transducin-like enhancer of split protein; cell fate;
KW differentiation; cervix cancer; breast cancer; psoriasis; baldness;
KW ss.
XX Homo sapiens.
XX Key Location/Qualifiers
FT CDS 26..2257
FT /*tag= a
XX
XX WO9407522-A.
XX
XX 14-APR-1994.
PD
XX 30-SEP-1993; 93WO-US09339.
PF
XX 30-SEP-1992; 92US-0954813.
PR
XX (UYVA ) UNIV YALE.
PA
XX Artavanis-tsakonas S, Stifani S;
PI
XX WPI; 1994-135221/16.
DR P-PSDB; AAR51110.
XX
XX Transducin-like enhancer or split proteins and nucleic acids-
PT are for treatment of disorders of cell fate or differentiation
PT e.g. cervical cancer, breast cancer, psoriasis, baldness etc.
XX
XX Disclosure: Page 91-94; 147pp; English.
XX
XX The nucleotide and deduced aa sequences of human transducin-like
CC enhancer of split proteins TLE-1 (AAQ45333, AAR51109), TLE-2 (AAQ45334,
CC AAR51110), TLE-3 (AAQ45335, AAR51111) and TLE-4 (AAQ45336, AAR52553)
CC were determined. The aa sequences were compared with that of
CC Drosophila E(spl) m9/10 (AAR52955). Comparison of the WD-40 domains
CC of these proteins defined the consensus residues shown in AAR92954.
CC The CcN motifs of the proteins were compared with those of SV40 T
CC antigen, human c-myc, human p53, human A-myc and dorsal protein with
CC respect to nuclear localization site, and casein-kinase and cd2-
CC kinase phosphorylation sites (sequences AAR52556-70).
XX
XX Sequence 2271 BP; 437 A; 751 C; 672 G; 411 T; 0 other;
SQ
Query Match 3.0%; Score 18; DB 15; Length 2271;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 16 tggagctgtgtcccgatg 33
DB 405 TGGAGCTGCTGCCCGATG 388
RESULT 16
AAA07836
ID AAA07836 standard; CDNA; 7746 BP.
XX
XX AAA07836;
AC
XX 07-JUL-2000 (first entry)
DT
XX Human homologue of UNC-53 (Hs-UNC-53/2) nucleotide sequence.
XX
XX UNC-53; Caenorhabditis elegans; microtubule; neural regeneration;
KW anticancer; anti-neurodegeneration; antifibrotic; anti-adhesive; human;
KW antisclerotic; antimeastatic; anti-arthritis; autoimmune disease; ss.
XX
XX Homo sapiens.
OS
XX Key Location/Qualifiers
FH

```

```

FI misc_feature 1..366
FI /*tag= a
FI /*note= "this region can be replaced with one of the
FI three sequences shown in AAA07837 to AAA07839;
FI this provides three variants for the 5' end
FI of the gene."
FI 5425..5433
FI misc_feature
FI /*tag= b
FI /*note= "this region is found to be absent in a
FI variant cDNA from Hela and colorectal
FI adenocarcinoma tissue."
FI 5924..6024
FI misc_feature
FI /*tag= c
FI /*note= "this region is found to be absent in a
FI variant cDNA isolated from frontal cortex;
FI Absence of this fragment results in an out-of-
FI frame deletion of 101 basepairs, resulting in
FI premature stop in translation."
XX
XX W09563080-A1.
XX
XX 05-DEC-1999.
XX
XX 02-JUN-1999; 99WO-EP03848.
XX
XX 03-JUN-1998; 98GB-OJ11962.
XX
XX (JANC ) JANSSEN PHARM NV.
XX
XX Luyten WHML, De Raeymaeker MC, Geysen JJGH, Boggaert TAOE;
XX Maerten LJS, Verhasselt P, Van De Craen M;
PI
XX WPI; 2000-116370/10.
XX
XX P-PSDB; AAR55565.
DR
XX
XX Novel proteins and nucleic acids e.g. for treating neurodegeneration
XX
XX Claim 94; Fig 1c; 146pp; English.
XX
XX The invention provides vertebrate (human) protein homologue of a UNC-53
XX protein of Caenorhabditis elegans. The UNC-53 binds to microtubules or
XX their plus ends. The UNC-53 sequences are used to promote neural
XX regeneration, revascularization and wound healing; also for treating
XX neurodegenerative disease, acute traumatic injury, fibrotic disease and
XX autoimmune diseases (e.g. rheumatoid arthritis and sclerosis). The UNC-53
XX polynucleotides can be used for recombinant production of the proteins,
XX as a source of probes for detecting allelic variants and polymorphisms,
XX for sequencing genomic DNA and for detecting UNC-53 expression; and as
XX source of therapeutic antisense sequences. Cells that express the
XX protein are used to identify regulators of cell shape, growth, motility
XX and migration. They can also be used to identify proteins that are
XX involved in signal transduction pathways also involving UNC-53, and to
XX identify compounds that alter attachment of UNC-53 to microtubules. A
XX target gene coupled to a UNC-53 encoding sequence may be used to deliver
XX the target gene to a cellular microtubule or its plus ends. The present
XX sequence represents the nucleotide sequence of the second human homologue
XX of UNC-53, designated Hs-UNC-53/2.
XX
XX Sequence 7746 BP; 1563 A; 2335 C; 2010 G; 1410 T; 8 other;
SQ
Query Match 3.0%; Score 18; DB 21; Length 7746;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 284 gctggagactctgtcccg 301
DB 1386 gctggagactctgtcccg 1403
RESULT 17
AAA0583
ID AAX40983 standard; cDNA; 298 BP.

```

```

XX AC AAX40983;
XX DT 18-JUN-1999 (first entry)
XX DE Human secreted protein 5' EST SEQ ID NO: 195.
XX KW Human: secreted protein; EST: expressed sequence tag; diagnosis;
XX KW forensic; gene therapy; chromosome mapping; signal peptide;
XX KW upstream regulatory sequence; cytokine activity; cell proliferation;
XX KW differentiation; haematopoiesis regulation; tissue growth regulation;
XX KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
XX KW thrombolytic; anti-inflammatory; tumour inhibition; ds.
XX OS Homo sapiens.
XX PN WO9906554-A2.
XX PD 11-FEB-1999.
XX PF 31-JUL-1998; 98WO-IB01238.
XX PR 01-AUG-1997; 97US-0905134.
XX PA (GEST ) GENSET.
XX PI Duclert A, Dumas Milne Edwards J, Lacroix B;
XX DR WPI: 1999-153784/13.
XX DR P-PSDB; AAY12150.
XX PT New nucleic acids encoding human secreted proteins - obtained from
XX PT cDNA libraries prepared from kidney, fetal kidney, dystrophic
XX PT muscle, muscle and heart tissue
XX PS Claim 1; Page 292; 622pp; English.
XX CC AAX40826 to AAX41093 represent 5' expressed sequence tags (ESTs) for
XX CC human secreted proteins, and encode the proteins given in AAY01602 and
XX CC AAY11994 to AAY12260, respectively. The proteins given represent the
XX CC signal peptide and an N-terminal fragment of a secreted protein. The
XX CC nucleic acid sequences can be used for producing secreted human gene
XX CC products. They can also be used to develop products for diagnosis and
XX CC therapy. The proteins obtained may have cytokine activity, cell
XX CC proliferation/differentiation activity, haematopoiesis regulating
XX CC activity, tissue growth regulating activity, reproductive hormone
XX CC regulating activity, chemotactic/chemokinetic activity, haemostatic and
XX CC thrombolytic activity, receptor/ligand activity, anti-inflammatory
XX CC activity, tumour inhibition activity or other activities. The products
XX CC can be used in forensic, gene therapy and chromosome mapping procedures.
XX CC The sequences can also be used for obtaining corresponding promoter
XX CC sequences. The nucleic acids encoding the signal peptide can be used
XX CC for directing extracellular secretion of a polypeptide or the insertion
XX CC of a polypeptide into a membrane, or importing a polypeptide into
XX CC a cell.
XX SQ Sequence 298 BP; 73 A; 66 C; 101 G; 58 T; 0 other;

Query Match 2.8%; Score 17; DB 20; Length 298;
Best Local Similarity 100.0%; Pred. No. 96;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 225 cgggcggctgctgaaga 241
DB 125 cgggcggctgctgaaga 141

RESULT 18
AAC06962/c
ID AAC06962 standard; cDNA: 385 BP.
XX
XX AC AAC06962;

```

```

XX 06-OCT-2000 (first entry)
XX DT 18-JUN-1999 (first entry)
XX DE Human secreted protein 5' EST, SEQ ID NO: 11037.
XX KW Human: 5' EST: expressed sequence tag; secreted protein; cDNA isolation;
XX KW gene therapy; chromosome mapping; ss.
XX OS Homo sapiens.
XX PN EP1033401-A2.
XX PD 06-SEP-2000.
XX PF 21-FEB-2000; 2000EP-0-000610.
XX PR 26-FEB-1999; 99US-0122487.
XX PA (GEST ) GENSET.
XX PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX DR WPI: 2000-500381/45.
XX PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
XX PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
XX PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX PS Claim 1; SEQ ID 11037; 71pp + CD-ROM; English.
XX CC The present sequence is one of a large number of 5' ESTs derived from
XX CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
XX CC identified within the present sequence. The 5' ESTs were prepared from
XX CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
XX CC sequences usually correspond mainly to the 3' untranslated region (UTR)
XX CC of the mRNA because they are often obtained from oligo-dT primed cDNA
XX CC libraries. Such ESTs are not well suited for isolating cDNA sequences
XX CC derived from the 5' ends of mRNAs and even in those cases where longer
XX CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
XX CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
XX CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
XX CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
XX CC They are used to obtain upstream regulatory sequences and to design
XX CC expression and secretion vectors.
XX SQ Sequence 385 BP; 94 A; 83 C; 122 G; 84 T; 2 other;

Query Match 2.8%; Score 17; DB 21; Length 385;
Best Local Similarity 100.0%; Pred. No. 95;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 477 aaggatgcccaagga 493
DB 185 AAGGATGCCAAGAGGA 169

RESULT 19
AAC06983/c
ID AAC06983 standard; cDNA: 413 BP.
XX
XX AC AAC06983;
XX DT 06-OCT-2000 (first entry)
XX DE Human secreted protein 5' EST, SEQ ID NO: 11058.
XX KW Human: 5' EST: expressed sequence tag; secreted protein; cDNA isolation;
XX KW gene therapy; chromosome mapping; ss.
XX OS Homo sapiens.
XX PN EP1033401-A2.

```

XX 06-SEP-2000.  
 XX 21-FEB-2000; 2000EP-0200610.  
 XX 26-FEB-1999; 990US-0122487.  
 XX (GSET ) GENSET.  
 XX Dumas Milne Edwards J, Duclert A, Giordano J;  
 XX WPI; 2000-500381/45.  
 XX  
 XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
 PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for  
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -  
 XX  
 XX Claim 1; SEQ ID 11058; 71pp + CD-ROM; English.  
 XX  
 XX The present sequence is one of a large number of 5' ESTs derived from  
 CC mRNAs encoding secreted proteins. No ORF has yet been conclusively  
 CC identified within the present sequence. The 5' ESTs were prepared from  
 CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST  
 CC sequences usually correspond mainly to the 3' untranslated region (UTR)  
 CC of the mRNA because they are often obtained from oligo-dT primed cDNA  
 CC libraries. Such ESTs are not well suited for isolating cDNA sequences  
 CC derived from the 5' ends of mRNAs and even in those cases where longer  
 CC cDNA sequences have been obtained, the full 5' UTR is rarely included.  
 CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be  
 CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used  
 CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.  
 CC They are used to obtain upstream regulatory sequences and to design  
 CC expression and secretion vectors.  
 XX  
 XX Sequence 413 BP; 95 A; 103 C; 115 G; 100 T; 0 other;  
 SQ

Query Match 2.8%; Score 17; DB 21; Length 413;  
 Best Local Similarity 100.0%; Pred. No. 95;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 477 aagatgccaagaagga 493  
 |||||  
 DB 322 AAGGATGCCAAGAGGA 306

RESULT 20  
 AAC76172  
 ID AAC76172 standard; cDNA; 477 BP.  
 XX  
 XX AAC76172;  
 XX  
 XX 08-FEB-2001 (first entry)  
 DT  
 DE Human ORFX ORF1727 polynucleotide sequence SEQ ID NO:3453.  
 XX  
 XX Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;  
 KW vulnary; antiparathic; antiparathic; antiparathic; antiparathic;  
 KW anticonvulsant; osteopathic; antiparathic; antiparathic; antiparathic;  
 KW immunostimulant; thrombolytic; coagulant; vasotrophic; antidiabetic;  
 KW hypotensive; dermatological; immunosuppressive; antihypertensive;  
 KW antiviral; antibacterial; antifungal; antiparathic; antiparathic;  
 KW neurodegenerative disorder; cancer; proliferative disorder; hypertension;  
 KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;  
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;  
 KW cholesterol ester storage; systemic lupus erythematosus; infection;  
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;  
 KW bone damage; cartilage damage; antihypertensive disease; coagulation;  
 KW thrombosis; contraceptive; ss.  
 XX  
 XX Homo sapiens.  
 XX

XX WO2000058473-A2.  
 XX  
 XX 05-OCT-2000.  
 XX  
 XX 31-MAR-2000; 2000WO-US08621.  
 XX  
 XX 31-MAR-1999; 990US-0127607.  
 PR 02-APR-1999; 990US-0127636.  
 PR 05-APR-1999; 990US-0127728.  
 PR 30-MAR-2000; 2000US-0540763.  
 XX  
 XX (CURA-) CURAGEN CORP.  
 XX  
 XX Shimkels RA, Leach M;  
 XX WPI; 2000-602362/57.  
 DR P-PSDB; AAB41563.  
 XX  
 XX Novel nucleic acids and peptides derived from open reading frame X,  
 PT useful for treating e.g. cancers, proliferative disorders,  
 PT neurodegenerative disorders and cardiovascular disease -  
 XX  
 XX Claim 5; Page 2627; 5507pp; English.  
 XX  
 XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,  
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX  
 CC sequences have activities such as: cytostatic; hepatotropic; vulnary;  
 CC antiparathic; antiparathic; antiparathic; antiparathic;  
 CC osteopathic; anticonvulsant; antiparathic; antiparathic;  
 CC immunostimulant; cardant; thrombolytic; coagulant; vasotrophic;  
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;  
 CC antihypertensive; antiparathic; antiparathic; antiparathic;  
 CC antihypertensive; antiparathic; antiparathic; antiparathic;  
 CC the presence of or predisposition to, or preventing or treating  
 CC pathological conditions associated with an ORFX-associated disorder. The  
 CC nucleic acids can be used to express ORFX proteins in gene therapy  
 CC vectors. The proteins and nucleic acids may be used to treat cancers,  
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,  
 CC graft vs host disease, cardiovascular disease, diabetes mellitus, lupus  
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,  
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,  
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,  
 CC nocturnal haemoglobinuria, antihypertensive disease; to enhance  
 CC coagulation; to inhibit thrombosis; and as a contraceptive.  
 XX  
 XX Sequence 477 BP; 113 A; 108 C; 162 G; 92 T; 2 other;  
 SQ

Query Match 2.8%; Score 17; DB 21; Length 477;  
 Best Local Similarity 100.0%; Pred. No. 95;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 225 cgggcggcgtcgtgaaga 241  
 |||||  
 DB 169 cgggcggcgtcgtgaaga 185

RESULT 21  
 AAA57874/C  
 ID AAA57874 standard; cDNA; 872 BP.  
 XX  
 XX AAA57874;  
 XX  
 XX 10-OCT-2000 (first entry)  
 DT  
 DE Leek glossy8 (g18) homologue partial cDNA, SEQ ID NO:5.  
 XX  
 XX Leek; glossy8 homologue; g18; cuticle; cuticular lipid biosynthesis;  
 KW lipid composition; lipid quantity; disease resistance; pest resistance;  
 KW wind resistance; frost resistance; UV resistance; transgenic plant;  
 KW antisense inhibition; ds.  
 XX  
 XX

OS Allium porrum.  
XX US060644-A.  
XX 05-MAY-2000.  
XX 25-DEC-1995; 95US-0581148.  
XX 24-MAR-1994; 94US-0218028.  
XX (IOWA ) UNIV IOWA STATE RES FOUND INC.  
XX Schnable PS, Nikolau BJ, Xu X, Xia Y, Robertson DS, Hansen JD;  
XX WPI; 2000-349707/30.  
XX Transforming plants such as maize and canola, for producing new plant  
PT varieties having disease and pest resistance involves introducing  
PT cuticular lipid genes into the plant genome .  
XX Claim 1; Column 25-26; 69pp; English.  
XX The invention relates to transforming a plant cell with either a nucleic  
CC acid encoding a plant cuticular lipid biosynthetic protein, or an  
CC antisense nucleic acid sequence targeted to a plant cuticular lipid  
CC gene, and then generating a plant from the plant cell. The cuticular  
CC lipid genes that may be used in the invention are given in  
CC AAA57871-A57884 and AAA57891. The cuticle consists of a meshwork (cutin)  
CC of cross- esterified polymerised hydroxy-fatty acids embedded in a  
CC complex mixture of nonpolar lipids (the cuticular wax). The cuticle  
CC protects and strengthens the plants, prevents evaporation of internal  
CC fluids and filters UV radiation. Mutations in the cuticular lipid genes  
CC affects the quantity and composition of cuticular lipids; in maize, 17  
CC loci (the glossy (gl) genes) have been identified as being involved in  
CC cuticular lipid biosynthesis or control. The method is useful for  
CC introducing a nucleic acid into a cell preferably of maize, soybean,  
CC rapeseed, canola, cotton, safflower, peanut, palm or sunflower plant and  
CC generating a plant from the plant cell. The transformed plants have  
CC improved environmental resistance relating to wind, frost, UV or drought;  
CC fungal, bacterial or viral disease resistance; pest resistance; and  
CC altered lipid content in seeds. Sequences AAA57874-A57875 represent leek  
CC glossy8 (gl8) homologue partial cDNA sequences.  
XX Sequence 872 BP; 269 A; 219 C; 165 G; 219 T; 0 other;  
Query Match 2.8%; Score 17; DB 21; Length 872;  
Best Local Similarity 100.0%; Pred. No. 93;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 475 tcaaggatgccaaagaag 491  
DB 209 TCAAGGATGCCAAGAAG 193  
RESULT 22  
AAA57875  
ID AAA57875 standard; cDNA; 872 BP.  
XX AAA57875;  
XX 10-OCT-2000 (first entry)  
XX Leek glossy8 (gl8) homologue partial cDNA, SEQ ID NO:6.  
XX Leek; glossy8 homologue; gl8; cuticle; cuticular lipid biosynthesis;  
KW lipid composition; lipid quantity; disease resistance; pest resistance;  
KW wind resistance; frost resistance; UV resistance; transgenic plant;  
KW antisense inhibition; ds.  
XX Allium porrum.  
OS Allium porrum.  
XX Key Location/Qualifiers

CDs 3..602  
FI /tag- a  
FI /partial  
FI /product= "Leek glossy8 homologue protein fragment"  
FI /note= "No initiation or termination codon given in the  
FI specification"  
FI /transl\_except= (pos:282..285, aa:Gly)  
FI /note= "There is an apparent 1 base insertion which  
FI alters the reading frame"  
FI /transl\_except= (pos:463..464, aa:Ala)  
FI /note= "There is an apparent 1 base deletion which  
FI alters the reading frame"  
FI misc\_feature 603..872  
FI /\*tag- C  
FI /note= "The amino acids encoded by this region are not  
FI included in the corresponding protein"  
XX US060644-A.  
XX 05-MAY-2000.  
XX 25-DEC-1995; 95US-0581148.  
XX 24-MAR-1994; 94US-0218028.  
XX (IOWA ) UNIV IOWA STATE RES FOUND INC.  
XX Schnable PS, Nikolau BJ, Xu X, Xia Y, Robertson DS, Hansen JD;  
XX WPI; 2000-349707/30.  
XX P-PSGB; AAB03103.  
XX Transforming plants such as maize and canola, for producing new plant  
PT varieties having disease and pest resistance involves introducing  
PT cuticular lipid genes into the plant genome .  
XX Claim 1; Column 27-28; 69pp; English.  
XX The invention relates to transforming a plant cell with either a nucleic  
CC acid encoding a plant cuticular lipid biosynthetic protein, or an  
CC antisense nucleic acid sequence targeted to a plant cuticular lipid  
CC gene, and then generating a plant from the plant cell. The cuticular  
CC lipid genes that may be used in the invention are given in  
CC AAA57871-A57884 and AAA57891. The cuticle consists of a meshwork (cutin)  
CC of cross- esterified polymerised hydroxy-fatty acids embedded in a  
CC complex mixture of nonpolar lipids (the cuticular wax). The cuticle  
CC protects and strengthens the plants, prevents evaporation of internal  
CC fluids and filters UV radiation. Mutations in the cuticular lipid genes  
CC affects the quantity and composition of cuticular lipids; in maize, 17  
CC loci (the glossy (gl) genes) have been identified as being involved in  
CC cuticular lipid biosynthesis or control. The method is useful for  
CC introducing a nucleic acid into a cell preferably of maize, soybean,  
CC rapeseed, canola, cotton, safflower, peanut, palm or sunflower plant and  
CC generating a plant from the plant cell. The transformed plants have  
CC improved environmental resistance relating to wind, frost, UV or drought;  
CC fungal, bacterial or viral disease resistance; pest resistance; and  
CC altered lipid content in seeds. Sequences AAA57874-A57875 represent leek  
CC glossy8 (gl8) homologue partial cDNA sequences.  
XX Sequence 872 BP; 219 A; 165 C; 218 G; 270 T; 0 other;  
Query Match 2.8%; Score 17; DB 21; Length 872;  
Best Local Similarity 100.0%; Pred. No. 93;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 475 tcaaggatgccaaagaag 491  
DB 664 tcaaggatgccaaagaag 680  
RESULT 23  
AAH50561/c

us-09-248-178-60.rng

Tue Apr 30 14:18:49 2002

XX	AAH50961 standard; cDNA; 991 BP.	XX	21-AUG-2000 (first entry)	XX	Human G protein-coupled receptor GPR17 cDNA.	XX	2.8%; Score 17; DB 21; Length 1020;
XX	AAH50961;	XX		XX	G protein-coupled receptor; GPCR; constitutively active;	XX	Best Local Similarity 100.0%; Pred. No. 93;
XX		XX		XX	intracellular loop 3; transmembrane domain 6; drug screening;	XX	Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX	28-AUG-2001 (first entry)	XX		XX	agonist; antagonist; ss.		
XX	Phosphatidylinositol synthase partial clone nucleotide sequence.	XX		XX	Homo sapiens.		
XX	Moss; Physcomitrella patens; lipid metabolism related protein; LMRP;	XX		XX	WO200022125-A1.		
XX	lipid biosynthesis; lipid modification; lipid degradation; cofactor;	XX		XX	20-APR-2000.		
XX	fatty acid transport; genetic engineering; fatty acid; enzyme; plant;	XX		XX	12-OCT-1999; 99WO-US23936.		
XX	microorganism; polyunsaturated fatty acid; oilseed plant; maize; wheat;	XX		XX	13-OCT-1998; 99US-0170496.		
XX	biotic stress tolerance; abiotic stress tolerance; rye; oat; triticale;	XX		XX	(AREN-) ARENA PHARM INC.		
XX	rice; barley; soybean; peanut; cotton; rapeseed; canola; manihot;	XX		XX	Behan DP; Chairlers DT; Liaw CM.		
XX	pepper; sunflower; tagetes; potato; tobacco; eggplant; tomato; Vicia;	XX		XX	WPI: 2000-32916/26.		
XX	pea; alfalfa; coffee; cacao; tea; Salix; oil palm; coconut;	XX		XX	P-PSDB; AAY50615.		
XX	perennial grass; forage crop; ss.	XX		XX	Non-endogenous constitutively activated human G protein-coupled		
XX	Physcomitrella patens.	XX		XX	receptors, useful for identifying agonists for use as pharmaceutical		
XX	WO200138484-A2.	XX		XX	agents		
XX	31-MAY-2001.	XX		XX	Example 1; Page 125-126; 34pp; English.		
XX	22-NOV-2000; 2000WO-EP11615.	XX		XX	The invention relates to constitutively active, non-endogenous versions		
XX	25-NOV-1999; 99WO-EP09108.	XX		XX	of endogenous human orphan G protein-coupled receptors (GPCRs, AAY50643-		
XX	(BADI) BASF PLANT SCI GMBH.	XX		XX	AAY50677 and AAY50683-Y50687), and to DNA encoding them (AAA30709-A30743		
XX	Lerchl J, Renz A, Ehrhardt T, Reindl A, Cirpus P, Bischoff F;	XX		XX	and AAA30775-A30779). The mutant proteins of the invention contain a		
XX	Frank M, Freund A, Duwenig E, Schmidt R, Reski R;	XX		XX	mutation in a portion of the protein comprising intracellular loop 3		
XX	WPI: 2001-367669/38.	XX		XX	(IC3) and transmembrane domain 6 (TM6). A non-endogenous amino acid, x,		
XX	Nucleic acids encoding lipid metabolism related proteins from	XX		XX	is substituted for an endogenous residue in IC3 at a position 16 amino		
XX	Physcomitrella patens useful to produce fine chemicals in modified	XX		XX	acids N-terminal. The endogenous amino acid is selected from Lys, His, Arg		
XX	organisms, particularly polyunsaturated fatty acids in oilseed plants -	XX		XX	x-(AA)15-Pro. The endogenous proline in TM6 to form a sequence		
XX	Claim 7; Page 105; 120pp; English.	XX		XX	or Ala, and is preferably Lys. When the endogenous residue at this		
XX	The present invention describes isolated nucleic acid sequences which	XX		XX	position is Lys, this residue is replaced by His, Arg or preferably Ala.		
XX	encode lipid metabolism related proteins (LMRP). The LMRP nucleic acids	XX		XX	The 15 amino acid stretch between the substituted amino acid and the pro		
XX	can be used to modify lipids and fatty acids, cofactors and enzymes in	XX		XX	may be endogenous, non-endogenous, or a mixture of endogenous and		
XX	microorganisms and plants, particularly to produce polyunsaturated fatty	XX		XX	non-endogenous residues. The constitutively active GPCRs are useful for		
XX	acids, and are especially useful in oilseed plants. The nucleic acids	XX		XX	identifying antagonists, agonists and partial agonists for use as		
XX	may also confer biotic or abiotic stress tolerance, particularly to	XX		XX	pharmaceutical agents. The mutant proteins are also useful in research		
XX	maize, wheat, rye, oat, triticale, rice, barley, soybean, peanut,	XX		XX	settings for elucidating the roles of the receptors in normal and		
XX	cotton, rapeseed, canola, manihot, pepper, sunflower, tagetes, potato,	XX		XX	diseased conditions. Antagonists for a particular GPCR are useful for		
XX	tobacco, eggplant, tomato, Vicia species, pea, alfalfa, coffee, cacao,	XX		XX	treating diseases and disorders associated with that receptor. Because		
XX	tea, Salix species, oil palm, coconut, perennial grasses and forage	XX		XX	the novel mutant GPCRs are constitutively active, they can be used		
XX	crops. AAH50878 to AAH50882 represent primers used in the exemplification	XX		XX	directly for screening of compounds without the need for endogenous		
XX	of the present invention. AAH50883 to AAH50968 represents LMRP nucleotide	XX		XX	ligands. The present sequence represents cDNA encoding a human wild-type		
XX	sequences, and AAG80843 to AAG80928 represent LMRP protein sequences,	XX		XX	GPCR used in an exemplification of the invention. This was cloned and		
XX	given in the present invention.	XX		XX	subjected to site-directed mutagenesis (SDM) to generate DNA encoding		
XX	Sequence 991 BP; 281 A; 214 C; 209 G; 287 T; 0 other;	XX		XX	the corresponding mutant of the invention.		
XX		XX		XX	Sequence 1020 BP; 174 A; 364 C; 261 G; 221 T; 0 other;		
XX	Query Match	XX		XX			
XX	Best Local Similarity 100.0%; Pred. No. 93;	XX		XX			
XX	Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	XX		XX			
XX		XX		XX			
XX	143 gccactccgaagctgc 159	XX		XX			
XX		XX		XX			
XX	240 GCCACTCCGAAAGCTGC 224	XX		XX			
XX		XX		XX			
XX	RESULT 24	XX		XX			
XX	AAA30601	XX		XX			
XX	ID AAA30601 standard; cDNA; 1020 BP.	XX		XX			
XX	AAA30601;	XX		XX			
XX		XX		XX			



DE Human secreted protein gene 12 SEQ ID NO:22.

XX Human; secreted protein; immunosuppressive; antiarthritic; antirheumatic;

KW antiproliferative; cytostatic; cardiant; vasotropic; cerebroprotective;

KW neuroprotective; antibacterial; virucide; fungicide;

KW ophthalmological; vulnary; autoimmune disease; rheumatoid arthritis;

KW hyperproliferative disorders; cancer; cardiovascular disorder;

KW cardiac arrest; cerebrovascular disorder; nervous system disorder;

KW Alzheimer's disease; ocular disorder; wound healing; skin aging; ss.

XX

OS Homo sapiens.

PN WO200061626-A1.

XX

PD 19-OCT-2000.

XX

PF 06-APR-2000; 2000WO-US09066.

XX

PR 09-APR-1999; 99US-0128698.

PR 20-JAN-2000; 2000US-0176926.

XX

PA (HUMA-) HUMAN GENOME SCI INC.

PA (ROSE/) ROSEN C A.

XX

PI Rosen CA, Ruben SM, Komatsoulis G;

XX

XX WPI: 2000-619227/59.

DR P-PSDB; AAB51838.

DR

XX New nucleic acid molecules encoding 49 human secreted proteins for

PT diagnosing, preventing or ameliorating medical conditions and used for

PT food additives or preservatives -

XX

PS Claim 1; Page 440; 516pp; English.

XX

CC Polynucleotide sequences AAC93479 - AAC93527 represent cDNA encoding

CC human secreted proteins AAB51827 - AAB51875. Sequences AAB51876 -

CC AAB51927 represent alternative polypeptides encoded by the genes, and

CC amino acid sequences with which they share homology. The genes and

CC proteins have activities dependent on the tissues and cells in which they

CC are expressed. Examples of their activities include immunosuppressive;

CC antiarthritic; antirheumatic; antiproliferative; cytostatic; cardiant;

CC vasotropic; cerebroprotective; neurotropic; neuroprotective; antibacterial;

CC virucide; fungicide; ophthalmological; and vulnary. The secreted

CC proteins, polynucleotides, antagonists and agonists may be useful in

CC treating, preventing and/or diagnosing diseases and disorders such as

CC autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative

CC disorders e.g. neoplasms of the breast or liver, cardiovascular disorders

CC e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,

CC angiogenesis, nervous system disorders e.g. Alzheimer's disease,

CC infections caused by bacteria, viruses and fungi and ocular disorders

CC e.g. corneal infection. The polypeptides can also be used to aid wound

CC healing and epithelial cell proliferation, to prevent skin aging due to

CC sunburn, to maintain organs before transplantation, for supporting cell

CC culture of primary tissues, to regenerate tissues and in chemotaxis. The

CC polypeptides can also be used as a food additive or preservative to

CC increase or decrease storage capabilities, fat content, lipid, protein,

CC carbohydrate, vitamins, minerals, cofactors and other nutritional

CC components. Oligonucleotides AAC93470 - AAC93478 and peptide AAB51826 are

CC used in the isolation and characterisation of the proteins and

CC polynucleotides of the invention.

XX

SQ Sequence 1258 BP; 291 A; 313 C; 330 G; 306 T; 18 other;

RESULT 28

AA066178

ID AA066178 standard; LNA; 1901 BP.

XX

AC AA066178.

XX

DI 07-FEB-1995 (first entry)

XX

DE Seven transmembrane receptor (R12) coding sequence.

XX

KW Primer; seven transmembrane receptor; receptor; amplification; ss.

XX

OS Homo sapiens.

XX

PH Key Location/Qualifiers

FT CUS 701..1720

FT /\*tag= a

FT /product= Seven transmembrane receptor (R12).

XX

PN W05412635-A.

XX

PD 09-JUN-1994.

XX

PF 17-NOV-1993; 93WO-US11153.

XX

PR 17-NOV-1992; 92US-0977452.

XX

PA (ICOS-) ICOS CORP.

XX

PI Godiska R, Gray PW, Schweickart VL;

XX

XX WPI: 1994-200264/24.

DR P-PSDB; AAR53753.

XX

XX DNA encoding seven transmembrane receptors - used to develop

PT prods. for use as therapeutic or diagnostic agents for conditions

PT involving the receptors.

XX

PS Example 10; Page 79-80; 100pp; English.

XX

CC PCR using two primers (AA066174, AA066175) was performed to amplify a

CC partial sequence of the R20 seven transmembrane coding sequence

CC which was later used as a probe for isolating the R20 genomic clone

CC (AA066176) from a human placenta DNA library. During the isolation

CC of the R20 gene two weakly hybridising sequences were identified

CC which had significant homology to other seven transmembrane

CC receptors. The probe was used to screen a human genomic foetal

CC liver DNA library, and while the R20 gene could not be identified in

CC this library, several weakly hybridising clones were plaque

CC purified, subcloned and sequenced. The two clones were designated

CC R2 (AA066177) and R12. This is the coding sequence of the R12 clone.

XX

SQ Sequence 1901 BP; 351 A; 640 C; 508 G; 372 T; 0 other;

Query Match 2.8%; Score 17; DB 15; Length 1901;

Best Local Similarity 100.0%; Pred. No. 51;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 323 caccagcgcgcctcgg 339

DB 1510 caccagcgcgcctcgg 1526

RESULT 29

AAV18356

ID AAV18356 standard; LNA; 1901 BP.

XX

AC AAV18356;

XX

DI 25-SEP-1998 (first entry)

XX

DR WPI; 1986-340201/52.

DR WPI: 1986-340201/52.

XX



DR P-PSDB: AAP60130.  
XX New human interleukin-1 polypeptide(s) - prepd. by recombinant  
PT DNA procedures for use in stimulating host immune systems, to  
PT promote wound healing etc.  
XX  
XX Disclosure: Fig 1: 37pp: English.  
XX Murine IL-1 encoding cDNA was used to isolate the structure of a  
CC partial genomic clone of human IL-1. IL-1 has utility in stimulating  
CC the immune system, especially for the promotion of wound healing, as  
CC a vaccine adjuvant, and in protection against neoplastic disease.  
CC IL-1 is also useful in treating critically ill protein-malnourished  
CC patients and in diagnostic assay systems.  
XX  
XX Sequence 1985 BP: 624 A; 436 C; 364 G; 561 T; 0 other;  
SQ  
Query Match 2.8%; Score 17; DB 7; Length 1985;  
Best Local Similarity 100.0%; Pred. No. 91;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 265 ataacgacgtgatgtca 281  
IIIIIIIIIIIIIIIIIIII  
Db 786 ATAAGCAGCTGATGTCA 770  
RESULT 32  
AAH14321/c  
ID AAH14321 standard; cDNA: 2040 BP.  
XX  
AC AAH14321;  
XX  
DT 26-JUN-2001 (first entry)  
XX  
DE Human cDNA sequence SEQ ID NO:11685.  
XX  
KW Human; primer: detection; diagnosis; antisense therapy; gene therapy; ss.  
XX  
OS Homo sapiens.  
XX  
PN EP1074617-A2.  
XX  
PD 07-FEB-2001.  
XX  
PF 28-JUL-2000; 2000EP-0116126.  
XX  
PR 29-JUL-1999; 99JP-0248036.  
PR 27-AUG-1999; 99JP-0300253.  
PR 11-JAN-2000; 2000JP-0118776.  
PR 02-MAY-2000; 2000JP-0183767.  
PR 09-JUN-2000; 2000JP-0241899.  
XX  
PA (HELI:) HELIX RES INST.  
XX  
XX Ora T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
XX  
XX WPI: 2001-318749/34.  
XX  
XX Primer sets for synthesizing polynucleotides, particularly the 5602  
PT full-length cDNAs defined in the specification, and for the detection  
PT and/or diagnosis of the abnormality of the proteins encoded by the  
PT full-length cDNAs -  
XX  
XX Claim 8; SEQ ID 11685; 2537pp + CD ROM; English.  
PS  
XX The present invention describes primer sets for synthesising 5602  
CC full-length cDNAs defined in the specification. Where a primer set  
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
CC to the complementary strand of a polynucleotide which comprises one of  
CC the 5602 nucleotide sequences defined in the specification, where the  
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination

CC of an oligonucleotide comprising a sequence complementary to the  
CC complementary strand of a polynucleotide which comprises a 5'-end  
CC sequence and an oligonucleotide comprising a sequence complementary to a  
CC polynucleotide which comprises a 3'-end sequence, where the  
CC oligonucleotide comprises at least 15 nucleotides and the combination of  
CC the 5'-end sequence/3'-end sequence is selected from those defined in  
CC the specification. The primer sets can be used in antisense therapy and  
CC in gene therapy. The primers are useful for synthesising polynucleotides,  
CC particularly full-length cDNAs. The primers are also useful for the  
CC detection and/or diagnosis of the abnormality of the proteins encoded by  
CC the full-length cDNAs. The primers allow obtaining of the full-length  
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
CC represent oligonucleotides, all of which are used in the exemplification  
CC of the present invention.  
XX  
XX Sequence 2040 BP: 629 A; 474 C; 468 G; 469 T; 0 other;  
SQ  
Query Match 2.8%; Score 17; DB 22; Length 2040;  
Best Local Similarity 100.0%; Pred. No. 91;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 401 ggaagcttcagaagaaa 417  
IIIIIIIIIIIIIIIIIIII  
Db 1160 GGAGCTTCAGAGGAAA 1144  
RESULT 33  
AAH144052  
ID AAH144052 standard; cDNA: 2453 BP.  
XX  
AC AAH144052;  
XX  
DT 28-FEB-1997 (first entry)  
XX  
DE Human G-protein thrombin-like receptor cDNA.  
XX  
KW G-protein; thrombin; receptor; diagnosis; gene therapy;  
KW haemophilia; wound healing; restenosis; angina; inflammation; ss.  
XX  
OS Homo sapiens.  
XX  
XX Key Location/Qualifiers  
FH CDS 548..1567  
FT /\*tag= a  
FT complement (548..565)  
FT /\*tag= b  
FT /\*note= \*5, primer binding site\*  
FT primer\_bind 1547..1564  
FT /\*tag= C  
FT /\*note= \*3, primer binding site\*  
XX  
XX WO9639438-A1.  
XX  
XX 12-DEC-1996.  
PD  
XX  
XX 06-JUN-1995; 95WO-US07180.  
XX  
XX 06-JUN-1995; 95WO-US07180.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Gocayne JD, Li Y, Ruben SM;  
PI  
XX WPI: 1997-043073/04.  
XX  
XX P-PSDB: AAW07617.  
XX  
XX Human G-protein thrombin receptor, HIBEB69 - useful to identify  
PI (ant)agonists, for treatment of angina, restenosis, wound healing  
PI etc.  
XX

PS Claim 1; Fig 1A-D; 58pp; English.

XX A cDNA clone (AA744092) codes for a human 7-transmembrane receptor  
 CC (AAW07617) putatively identified as a G-protein thrombin receptor. It  
 CC was discovered in a human infant brain cDNA library. G-protein  
 CC thrombin-like receptor polynucleotides can be used to produce  
 CC recombinant receptor in host cells (e.g. E. coli, COS-7, Sf9), to  
 CC design probes for detecting diseases related to receptor gene  
 CC mutation, and in gene therapy to treat haemophilia or promote wound  
 CC healing; antisense sequences can be used to treat inflammation,  
 CC restenosis, angina, etc.

XX SQ Sequence 2453 BP; 494 A; 815 C; 640 G; 504 T; 0 other;

Query Match 2.8%; Score 17; DB 18; Length 2453;

Best Local Similarity 100.0%; Pred. No. 91;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 323 caccacgcgcctcctgg 339

Db 1357 caccacgcgcctcctgg 1373

RESULT 34

ID AAH66941 standard; DNA: 2766 BP.

XX AAH66941;

XX 26-SEP-2001 (first entry)

XX C glutamicum coding sequence fragment SEQ ID NO: 1976.

XX Corynebacterium bacterium; amino acid synthesis; vitamin; saccharide;

XX organic acid synthesis; ds.

XX Corynebacterium glutamicum.

XX EP1106790-A2.

XX 20-JUN-2001.

XX 18-DEC-2000; 2000EP-0127688.

XX 16-DEC-1999; 99JP-0377484.

XX 07-APR-2000; 2000JP-0159162.

XX 03-AUG-2000; 2000JP-0280986.

XX (KYOWA) KYOWA HAKKO KOGYO KK.

XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;

XX Tateishi N, Senoh A, Ikeda M, Ozaki A;

XX WPI: 2001-376931/40.

XX P-PSDB; AAG91722.

XX Novel polynucleotides derived from Corynebacterium bacteria, for identifying  
 PT mutation point of a gene, measuring expression of a gene, analysing  
 PT expression profile or pattern of a gene and identifying homologous gene

XX Claim 8; SEQ ID NO: 1976; 246pp + Sequence Listing; English.

XX The present invention provides a number of nucleotide and protein  
 CC sequences from the Corynebacterium bacterium Corynebacterium glutamicum. These  
 CC are useful for identifying the mutation point of a gene derived from a  
 CC mutant of corynebacterium bacterium, measuring expression amount and  
 CC analysing the expression profile or expression pattern of a gene derived  
 CC from Corynebacterium bacterium, and identifying a homologue of a gene derived  
 CC from corynebacterium bacterium. Corynebacterium bacteria are useful for producing  
 CC amino acids, nucleic acids, vitamins, saccharides and organic acids,  
 CC particularly L-lysine. The present sequence is a nucleic acid described

CC in the exemplification of the invention.

CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from the  
 CC European Patent Office.

XX SQ Sequence 2766 BP; 695 A; 610 C; 838 G; 623 T; 0 other;

Query Match 2.8%; Score 17; DB 22; Length 2766;

Best Local Similarity 100.0%; Pred. No. 90;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 102 ggtcagtgaggagagc 116

Db 1667 ggtcagtgaggagagc 1703

RESULT 35

ID AAC75611 standard; cDNA: 2813 BP.

XX AAC75611;

XX 06-FEB-2001 (first entry)

XX Human ORFX ORF1166 polynucleotide sequence SEQ ID NO:2331.

XX Human; open reading frame; ORFX; detection; cytosolic; hepatotropic;  
 KW vulnary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;  
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;  
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
 KW hypotensive; dermatological; immunosuppressive; antiinflammatory;  
 KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;  
 KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;  
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;  
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;  
 KW cholesterol ester storage; systemic lupus erythematosus; infection;  
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;  
 KW bone damage; cartilage damage; antiinflammatory disease; coagulation;  
 KW thrombosis; contraceptive; ss.

XX Homo sapiens.

XX WJ200058473-A2.

XX G5-G11-2000.

XX 31-MAR-2000; 2000WG-US06621.

XX 31-MAR-1999; 99US-0127607.

XX 02-APR-1999; 99US-0127636.

XX 05-APR-1999; 99US-0127728.

XX 30-MAR-2000; 2000US-0540763.

XX (CUKA-) CURAGEN CORP.

XX Shinketsu RA, Leach M;

XX WPI: 2000-602362/57.

XX P-PSDB; AAB41102.

XX Novel nucleic acids and peptides derived from open reading frame X,  
 PT useful for treating e.g. cancers, proliferative disorders,  
 PT neurodegenerative disorders and cardiovascular disease

PS Claim 5; Page 1704-1706; 5507pp; English.

XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,  
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX  
 CC sequences have activities such as: cytostatic; hepatotropic; vulnary;  
 CC antipsoriatic; antiparkinsonian; nootropic; neuroprotective;  
 CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;

CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;  
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;  
 CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;  
 CC antithyroid; and antianemic. The sequences can be used for determining  
 CC the presence of or predisposition to, or preventing or treating  
 CC pathological conditions associated with an ORFX-associated disorder. The  
 CC nucleic acids can be used to express ORFX proteins in gene therapy  
 CC vectors. The proteins and nucleic acids may be used to treat cancers,  
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,  
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,  
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus  
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,  
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,  
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,  
 CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance  
 CC coagulation; to inhibit thrombosis; and as a contraceptive.  
 XX  
 SQ Sequence 2813 BP; 847 A; 657 C; 637 G; 670 T; 2 other;

Query Match 2.8%; Score 17; DB 21; Length 2813;  
 Best Local Similarity 100.0%; Pred. No. 90;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 401 ggagcttcagaagaaa 417  
 |||||  
 DB 1919 GGAGCTTCAGAGGAAA 1903

RESULT 36  
 AAI60348/C  
 ID AAI60348 standard; cDNA; 3092 BP.

XX AA160348;  
 XX 22-OCT-2001 (first entry)  
 XX Human polynucleotide SEQ ID NO 4337.

XX Human; nontropic; immunosuppressant; cytostatic; gene therapy; cancer;  
 KW peripheral nervous system; neuropathy; central nervous system; CNS;  
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
 KW leukaemia; ss.

XX Homo sapiens.  
 XX WO200153312-A1.  
 XX 26-JUL-2001.  
 XX 26-DEC-2000; 2000WO-US34263.

XX 21-JAN-2000; 2000US-0488725.  
 PR 25-APR-2000; 2000US-0552317.  
 PR 09-JUL-2000; 2000US-0598042.  
 PR 19-JUL-2000; 2000US-0620312.  
 PR 03-AUG-2000; 2000US-0653450.  
 PR 14-SEP-2000; 2000US-0662191.  
 PR 19-OCT-2000; 2000US-0693036.  
 PR 29-NOV-2000; 2000US-0727344.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;  
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;

XX WPI; 2001-442253/47.

DR P-PSDB; AAM41192.

XX Novel nucleic acids and polypeptides, useful for treating disorders

PI such as central nervous system injuries -  
 XX  
 PS Claim 1: SEQ ID NO 4337; 10078pp; English.  
 XX  
 CC The invention relates to human nucleic acids (AA157758-AA161369) and  
 CC the encoded polypeptides (AAM38642-AAM412213) with nootropic,  
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful  
 CC in gene therapy. A composition containing a polypeptide or polynucleotide  
 CC of the invention may be used to treat diseases of the peripheral nervous  
 CC system, such as peripheral nervous injuries, peripheral neuropathy and  
 CC localised neuropathies and central nervous system diseases, such as  
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
 CC utilisation of the activities such as: Immune system suppression,  
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
 CC assays for receptor activity, arthritis and inflammation, leukaemia and  
 CC C.N.S disorders.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification.  
 XX  
 SQ Sequence 3092 BP; 921 A; 727 C; 698 G; 746 T; 0 other;

Query Match 2.8%; Score 17; DB 22; Length 3092;  
 Best Local Similarity 100.0%; Pred. No. 90;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 401 ggagcttcagaagaaa 417  
 |||||  
 DB 2182 GGAGCTTCAGAGGAAA 2166

RESULT 37  
 AAH90050/C  
 ID AAH90050 standard; cDNA; 3092 BP.

XX AAH90050;  
 XX 01-OCT-2001 (first entry)  
 XX Human bone marrow cDNA, SEQ ID NO: 294.

XX Human; bone marrow; antiinflammatory; cytostatic; neuroprotective;  
 KW antiviral; antibacterial; antifungal; anti-HIV; haemostatic;  
 KW immunosuppressive; gene therapy; cytokine cell proliferation;  
 KW cell differentiation modulator; immune disorder; infection; cancer;  
 KW human immunodeficiency virus; HIV; autoimmune disorder; haemophilia; ss.

XX Homo sapiens.  
 XX WO200153453-A2.  
 XX 26-JUL-2001.

XX 23-DEC-2000; 2000WO-US34560.  
 XX 21-JAN-2000; 2000US-0488725.  
 PR 25-APR-2000; 2000US-0552317.  
 PR 09-JUL-2000; 2000US-0598042.  
 PR 19-JUL-2000; 2000US-0620312.  
 PR 03-AUG-2000; 2000US-0653450.  
 PR 14-SEP-2000; 2000US-0662191.  
 PR 19-OCT-2000; 2000US-0693036.  
 PR 30-NOV-2000; 2000US-0725083.

XX (HYSE-) HYSEQ INC.

XX Ford JE, Boyle BJ, Tang YT, Liu C, Asundi V, Chen R, Ma Y;  
 PI Ren F, Wang J, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;  
 PI Zhao QA, Zhou P, Drmanac RT;

XX WPI; 2001-488707/53.

DR P-PSDB: AAM00931.  
 XX Novel bone-marrow-expressed polynucleotides and polypeptides, useful  
 PT for treating e.g. cancer and immune deficiency disorders -  
 XX  
 XX Claim 1; Page 405; 648pp; English.  
 PS  
 XX The present sequence is one of 251 novel human polynucleotides  
 CC expressed in the bone marrow. The polynucleotide and the  
 CC polypeptide encoded by it are useful in the treatment of various  
 CC immune deficiencies and disorders. The deficiencies and disorders may  
 CC be genetic, may be caused by a viral (e.g. HIV), bacterial or fungal  
 CC infection, or may result from an autoimmune disorder, a coagulation  
 CC disorder (e.g. haemophilia), inhibition of tumour cell proliferation,  
 CC suppression of an inflammatory response or treatment of a nervous  
 CC system disorder such as Alzheimer's disease. Detection of the presence  
 CC or increased expression of the polynucleotide or the protein it  
 CC encodes is useful for the diagnosis and/or prognosis of one  
 CC or more types of cancer. The polynucleotide and polypeptide can be  
 CC used as nutritional sources or supplements and in the screening of  
 CC chemical compounds as potential drugs.  
 XX Sequence 3092 BP; 921 A; 727 C; 698 G; 746 T; 0 other;  
 SQ

Query Match 2.8%; Score 17; DB 22; Length 3092;  
 Best Local Similarity 100.0%; Pred. No. 90;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 401 ggagcttcagaaggaaa 417  
 |||||  
 Db 2182 GGAGCTTCAGAGGAAA 2166

RESULT 38  
 AAI58562/C  
 ID AAI58562 standard; cDNA: 3382 BP.  
 XX  
 AC. AAI58562;  
 XX  
 XX 22-OCT-2001 (first entry)  
 DT  
 DE Human polynucleotide SEQ ID NO 765.  
 KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;  
 KW peripheral nervous system; neuropathy; central nervous system; CNS;  
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
 KW leukaemia; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 XX WO200153312-A1.  
 PN  
 PD 26-JUL-2001.  
 XX  
 XX 26-DEC-2000; 2000WO-US34263.  
 XX  
 XX 21-JAN-2000; 2000US-0488725.  
 PR 25-APR-2000; 2000US-0552317.  
 PR 09-JUL-2000; 2000US-0598042.  
 PR 19-JUL-2000; 2000US-0620312.  
 PR 03-AUG-2000; 2000US-0653450.  
 PR 14-SEP-2000; 2000US-0662191.  
 PR 19-OCT-2000; 2000US-0693036.  
 PR 29-NOV-2000; 2000US-0727344.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;  
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;

XX WPI: 2001-442253/47.  
 DR P-PSDB; AAM39406.  
 XX Novel nucleic acids and polypeptides, useful for treating disorders  
 PT such as central nervous system injuries -  
 XX  
 XX Claim 1; SEQ ID NO 765; 10078pp; English.  
 PS  
 XX The invention relates to human nucleic acids (AAI57798-AAI61369) and  
 CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,  
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful  
 CC in gene therapy. A composition containing a polypeptide or polynucleotide  
 CC of the invention may be used to treat diseases of the peripheral nervous  
 CC system, such as peripheral nervous injuries, peripheral neuropathy and  
 CC localised neuropathies and central nervous system diseases, such as  
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
 CC utilisation of the activities such as: immune system suppression,  
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
 CC assays for receptor activity, arthritis and inflammation, leukaemias and  
 CC C.N.S disorders.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC Specification.  
 XX  
 SQ Sequence 3382 BP; 573 A; 797 C; 763 G; 849 T; 0 other;

Query Match 2.8%; Score 17; DB 22; Length 3382;  
 Best Local Similarity 100.0%; Pred. No. 90;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 401 ggagcttcagaaggaaa 417  
 |||||  
 Db 2450 GGAGCTTCAGAGGAAA 2474

RESULT 39  
 AAH89937/C  
 ID AAH89937 standard; cDNA: 3405 BP.  
 XX  
 AC. AAH89937;  
 XX  
 XX 01-OCT-2001 (first entry)  
 DT  
 DE Human bone marrow cDNA, SEQ ID NO: 68.  
 XX  
 KW Human; bone marrow; antiinflammatory; cytostatic; neuroprotective;  
 KW antiviral; antibacterial; antifungal; anti-HIV; haemostatic;  
 KW immunosuppressive; gene therapy; cytokine cell proliferation;  
 KW cell differentiation modulator; immune disorder; infection; cancer;  
 KW human immunodeficiency virus; HIV; autoimmune disorder; haemophilia; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 XX WO200153453-A2.  
 PN  
 PD 26-JUL-2001.  
 XX  
 XX 23-DEC-2000; 2000WO-US34960.  
 XX  
 XX 21-JAN-2000; 2000US-0488725.  
 PR 25-APR-2000; 2000US-0552317.  
 PR 09-JUL-2000; 2000US-0598042.  
 PR 19-JUL-2000; 2000US-0620312.  
 PR 03-AUG-2000; 2000US-0653450.  
 PR 14-SEP-2000; 2000US-0662191.  
 PR 19-OCT-2000; 2000US-0693036.  
 PR 30-NOV-2000; 2000US-0250583.  
 XX  
 PA (HYSE-) HYSEQ INC.

Tue Apr 30 14:18:49 2002

PI Ford JE, Boyle BJ, Tang YT, Liu C, Asundi V, Chen R, Ma Y;  
PI Ren F, Wang J, Werhman T, Xu C, Xue AJ, Yang Y, Zhang J;  
PI Zhao QA, Zhou P, Drmanac RT;  
XX  
XX WPI: 2001-488707/53.  
DR P-PSDB: AAM00818.  
XX  
XX Novel bone-marrow-expressed polynucleotides and polypeptides, useful  
PT for treating e.g. cancer and immune deficiency disorders -  
PI  
XX Claim 1: Page 252-253; 648pp; English.  
XX  
XX The present sequence is one of 251 novel human polynucleotides  
CC expressed in the bone marrow. The polynucleotide and the  
CC polypeptide encoded by it are useful in the treatment of various  
CC immune deficiencies and disorders. The deficiencies and disorders may  
CC be genetic, may be caused by a viral (e.g. HIV), bacterial or fungal  
CC infection, or may result from an autoimmune disorder, a coagulation  
CC disorder (e.g. haemophilia), inhibition of tumour cell proliferation,  
CC suppression of an inflammatory response or treatment of a nervous  
CC system disorder such as Alzheimer's disease. Detection of the presence  
CC or increased expression of the polynucleotide or the protein it  
CC encodes is useful for the diagnosis and/or prognosis of one  
CC or more types of cancer. The polynucleotide and polypeptide can be  
CC used as nutritional sources or supplements and in the screening of  
CC chemical compounds as potential drugs.  
XX  
SQ Sequence 3405 BP; 977 A; 803 C; 770 G; 855 T; 0 other;

Query Match 2.8%; Score 17; DB 22; Length 3405;  
Best Local Similarity 100.0%; Pred. No. 90;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 401 ggagcttcagaagaa 417  
|||||  
DB 2469 GGAGCTTCAGAGGAAA 2453

RESULT 40  
AAC76360/c  
ID AAC76360 standard; cDNA: 5713 BP.

XX  
AC AAC76360;

XX 08-FEB-2001 (first entry)

XX Human ORFX ORF1915 polynucleotide sequence SEQ ID NO:3829.

XX Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;  
KW vulnary; antiproliferative; antiparkinsonian; neurotropic; neuroprotective;  
KW immunostimulant; osteoplastic; antiarthritic; immunosuppressant; cardiac;  
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
KW hypotensive; dermatological; immunosuppressive; antinflammatory;  
KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;  
KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;  
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;  
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;  
KW cholesterol ester storage; systemic lupus erythematosus; infection;  
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;  
KW bone damage; cartilage damage; antinflammatory disease; coagulation;  
KW thrombosis; contraceptive; ss.

XX Homo sapiens.

OS WO200058473-A2.

XX 05-OCT-2000.

XX 31-MAR-2000; 2000WO-US08621.

XX 31-MAR-1999; 99US-0127607.

PR 02-APR-1999; 99US-0127636.  
PR 05-APR-1999; 99US-0127728.  
PR 30-MAR-2000; 2000US-0540763.  
XX  
XX (CURA-) CURAGEN CORP.  
XX  
XX Shimketa RA, Leach M;  
PI WPI: 2000-50262/57.  
DR P-PSUB: AAB42151.  
XX  
XX Novel nucleic acids and peptides derived from open reading frame X,  
PT useful for treating e.g. cancers, proliferative disorders,  
PI neurodegenerative disorders and cardiovascular disease -  
XX  
XX Claim 5: Page 2982-2985; 5507pp; English.

XX AAC7446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,  
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX  
CC sequences have activities such as: cytostatic; hepatotropic; vulnary;  
CC antiproliferative; antiparkinsonian; neurotropic; neuroprotective;  
CC osteoplastic; anticonvulsant; antiarthritic; immunosuppressant;  
CC immunostimulant; cardiac; thrombolytic; coagulant; vasotropic;  
CC antidiabetic; hypotensive; dermatological; immunosuppressive;  
CC antinflammatory; antibacterial; antiviral; antifungal; antirheumatic;  
CC antithyroid; and antianemic. The sequences can be used for determining  
CC the presence of or predisposition to, or preventing or treating  
CC pathological conditions associated with an ORFX-associated disorder. The  
CC nucleic acids can be used to express ORFX proteins in gene therapy  
CC vectors. The proteins and nucleic acids may be used to treat cancers,  
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,  
CC graft vs host disease, cardiovascular disease, diabetes mellitus,  
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus  
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,  
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,  
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,  
CC nocturnal haemoglobinuria, antinflammatory disease; to enhance  
CC coagulation; to inhibit thrombosis; and as a contraceptive.

XX Sequence 5713 BP; 1237 A; 1565 C; 1581 G; 1329 T; 1 other;

Query Match 2.8%; Score 17; DB 21; Length 5713;

Best Local Similarity 100.0%; Pred. No. 89;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 298 ccagagatgtgcccgg 314  
|||||

DB 3941 CCAGAGATGTGTCCGG 3925

RESULT 41  
AAH57490/c  
ID AAH57490 standard; cDNA: 5813 BP.

XX AAH57490;

XX 10-SEP-2001 (first entry)

XX Human liver cell specific cDNA sequence SEQ ID NO:330.

XX Human; tissue specific; diagnosis; brain; heart; skeletal muscle;  
KW lung; liver; uterus; ovary; stomach; intestine; kidney; pancreas; ss;  
KW metabolic disease; developmental disease; cytostatic; immunomodulatory;  
KW neuroprotective; gene therapy; cancer; immunopathology; neuropathology.

XX Homo sapiens.

OS WO200132527-A2.

XX 10-MAY-2001.

XX 02-NOV-2000; 2000WO-US30396.

```

XX 04-NOV-1999; 99US-0163508.
XX (INCY-) INCYTE GENOMICS INC.
XX Sornasse T, Selthamer JJ, Watson GA;
XX WPI; 2001-291057/30.
XX
XX New cell and tissue specific polynucleotides useful for diagnosis,
XX prognosis or monitoring of treatments for disorders where the gene is
XX associated with a cancer, immunopathology or neuropathology -
XX
XX Claim 1; Page 248-249; 327pp; English.
XX
XX AAH57161 to AAH57576 represent cell and tissue specific polynucleotide
XX sequences (I). (I) can have cytostatic, immunomodulatory and
XX neuroprotective activities, and can be used in gene therapy. (I) and
XX proteins (II) encoded by then are used in high throughput screening
XX assays to select DNA molecules, RNA molecules, peptide nucleic acids,
XX mimetics, peptides, proteins, agonists, antagonists, antibodies or
XX their fragments, immunoglobulins, inhibitors, drug compounds and
XX pharmaceutical agents. Expression of (I) in a sample indicates the
XX differentiation of embryonic stem cells into a tissue selected from
XX brain, heart, kidney, liver, lung, skeletal muscle or pancreatic
XX tissues. (I) and (II) are used to produce an expression profile that
XX defines a metabolic or developmental process, treatment, condition,
XX disease or disorder. The gene profile can be used for diagnosis,
XX prognosis or monitoring of treatments and for investigating a
XX predisposition to a disorder where the gene is associated with a
XX cancer, immunopathology or neuropathology.
XX
XX Sequence 5813 BP; 1290 A; 1665 C; 1665 G; 1193 T; 0 other:
XX
XX Query Match 2.8%; Score 17; DB 22; Length 5813;
XX Best Local Similarity 100.0%; Pred. No. 89;
XX Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 237 gaagagggccgcagga 253
DB 5426 GAAGAGGCCCGCAGGAA 5410
XX
RESULT 42
AAV52730
ID AAV52730 standard; DNA; 6441 BP.
XX
XX AAV52730;
XX
XX 21-DEC-1998 (first entry)
XX
XX Human hepatocyte nuclear factor 1 beta gene (TCF2).
XX
XX Hepatocyte nuclear factor 1 beta; HNF-1 beta; MODY4; human;
XX transcription factor; maturity onset diabetes of the young;
XX diabetes; NIDDM; diagnosis; therapy; ss.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX CDS 531..5556
XX /tag= a
XX /note= "contains introns"
XX
XX exon 356..874
XX /tag= b
XX /number= 1
XX intron 875..1312
XX /tag= c
XX /number= 1
XX misc_difference 1100
XX /tag= d
XX /note= "n at position 1100 represents an intronic

```

```

FI exon 1313..1512
FI /tag= e
FI /number= 2
FI intron 1513..2108
FI /tag= f
FI /number= 2
FI misc_difference 1750
FI /tag= g
FI /note= "n at position 1750 represents an intronic
FI exon 2109..2373
FI /tag= h
FI /number= 3
FI intron 2374..2592
FI /tag= i
FI /number= 3
FI misc_difference 2512
FI /tag= j
FI /note= "n at position 2512 represents an intronic
FI exon 2593..2828
FI /tag= k
FI /number= 4
FI intron 2829..3266
FI /tag= l
FI /number= 4
FI misc_difference 3003
FI /tag= m
FI /note= "n at position 3003 represents an intronic
FI exon 3267..3427
FI /tag= n
FI /number= 5
FI intron 3428..3790
FI /tag= o
FI /number= 5
FI misc_difference 3609
FI /tag= p
FI /note= "n at position 3609 represents an intronic
FI exon 3791..3923
FI /tag= q
FI /number= 6
FI intron 3924..4439
FI /tag= r
FI /number= 6
FI misc_difference 4185
FI /tag= s
FI /note= "n at position 4185 represents an intronic
FI exon 4440..4635
FI /tag= t
FI /number= 7
FI intron 4636..5143
FI /tag= u
FI /number= 7
FI misc_difference 4857
FI /tag= v
FI /note= "n at position 4857 represents an intronic
FI exon 5144..5262
FI /tag= w
FI /number= 8
FI intron 5263..5475
FI /tag= x
FI /number= 8
FI misc_difference 5369
FI /tag= y
FI /note= "n at position 5369 represents an intronic
FI exon 5476..6441
FI /tag= z

```

```

FT XX                               /number= 9
PN XX
PD XX WO9811254-A1.
PE XX
PF XX 19-MAR-1998.
PI XX
PR XX 10-SEP-1997; 97WO-US16037.
PR XX
PR XX 30-OCT-1996; 96US-0029679.
PR XX 10-SEP-1996; 96US-0025719.
PR XX 02-OCT-1996; 96US-0028056.
PA XX (ARCH-) ARCH DEV CORP.
XX
XX Bell GI, Furuta H, Horikawa Y, Kaisaki PJ, Menzel S;
PI Oda N, Yamagata K;
XX
XX WPI; 1998-271667/24.
DR P-PSDB; AAW71581 and AAW71586.
XX
XX Isolated nucleic acid encoding hepatocyte nuclear factor 1-alpha and
PT 1-beta - useful for detecting susceptibility for non-insulin
PT dependent diabetes, especially maturity-onset diabetes of the young
XX
XX Claim 89; Fig 27A-I; 363pp; English.
XX
XX This is a partial nucleotide sequence of the gene (TCF2) encoding
CC human hepatocyte nuclear factor 1 beta (HNF-1 beta, see AAW71581), a
CC homeodomain-containing transcription factor. Mutations in this
CC gene are indicative of a propensity to diabetes mellitus. The
CC invention concerns the identification of genes responsible for
CC non-insulin dependent diabetes mellitus (NIDDM) for use in
CC diagnostics and therapeutics. It demonstrates that the MODY3 locus
CC is the HNF-1 alpha gene (see AAV52625). It also demonstrates that the
CC beta gene (see also AAV52730) and the MODY4 locus is the HNF-4 alpha
CC gene (see AAV52687). Analysis of mutations in these HNF genes can be
CC diagnostic for diabetes. The invention also contemplates methods of
CC screening for modulators of HNF function utilising HNF nucleic acids
CC or polypeptides, the modulators being useful for treating diabetes
CC by modulating HNF function in an animal.
XX
SQ Sequence 6441 BP; 1572 A; 1827 C; 1530 G; 1504 T; 8 other;

Query Match      2.8%; Score 17; DB 19; Length 6441;
Best Local Similarity 100.0%; Pred. No. 89;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 245 ccgcaggaggaggagg 261
Db 2685 ccgcaggaggaggagg 2705
|||||

RESULT 43
AAV52207
ID AAV52207 standard; DNA; 16535 BP.
XX
XX AAV52207;
AC
XX
XX 23-OCT-1998 (first entry)
DT
XX
DE Streptococcus pneumoniae genome fragment SEQ ID NO:74.
XX
XX Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;
KW computer readable medium; vaccine; pharmaceutical composition; ds.
XX
XX Streptococcus pneumoniae.
OS
XX
XX WO9818931-A2.
PN
XX 07-MAY-1998.
PD
XX
XX 30-OCT-1997; 97WO-US19588.
PF

```

```

XX
PR 31-OCT-1996; 96US-0029560.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M;
PI Kunsch CA, Rusan CA;
XX
XX WPI; 1998-272225/24.
DR
XX
XX Computer-readable medium with recorded Streptococcus pneumoniae
PI polynucleotide sequences - useful in diagnostic kits and assays, and
PI pharmaceutical compositions and vaccines for Streptococcus
PI pneumoniae
XX
XX Claim 1; Page 608-617; 1409pp; English.
XX
XX The present invention describes a computer readable medium which has
CC the nucleotide sequences SEQ ID NO:1 to 391 (AAV52134 to AAV52524)
CC recorded on it, or a representative fragment or a sequence at least 95%
CC identical to SEQ ID NO:1 to 391. The nucleotide sequences depicted in
CC SEQ ID NO:1 to 391 (AAV52134 to AAV52524) are genomic fragments from
CC Streptococcus pneumoniae. The present invention also describes an
CC isolated nucleic acid molecule encoding a homologue of any of the
CC fragments of the S. pneumoniae genome (SEQ ID NO:1 to 391) where the
CC nucleic acid molecule is produced by a process comprising: (a) screening
CC a genomic DNA library using as a probe a target sequence defined by any
CC of the sequences in SEQ ID NO:1 to 391, identifying members of the
CC library which contain sequences that hybridise to the target sequence and
CC isolating the nucleic acid molecules from the members; or (b) isolating
CC mRNA, DNA or cDNA produced from an organism, amplifying nucleic acid
CC molecules whose nucleotide sequence is homologous to amplification
CC primers derived from the fragment of the S. pneumoniae genome to prime
CC the amplification and isolating the amplified sequences. The computer
CC readable medium can be used in a computer-based system for identifying
CC fragments of the S. pneumoniae genome of commercial importance, or
CC expression modulating fragments of the S. pneumoniae genome. Products
CC from the present invention can be used in diagnosis kits and assays, and
CC pharmaceutical compositions and vaccines for S. pneumoniae.
XX
XX Sequence 16535 BP; 5001 A; 3914 C; 3102 G; 4518 T; 0 other;
SQ

```

```

Query Match      2.8%; Score 17; DB 19; Length 16535;
Best Local Similarity 100.0%; Pred. No. 86;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 474 ttcaaggatgccaaaga 490
Db 451 ttcaaggatgccaaaga 467
|||||

RESULT 44
AAF88315
ID AAF88315 standard; DNA; 45624 BP.
XX
XX AAF88315;
AC
XX
XX 28-AUG-2001 (first entry)
DT
XX
XX S. spinosa DNA fragment SEQ ID 4.
DE
XX
XX Forosamin; trimethylrhamsone; polyketide synthase; biosynthesis;
KW spinosyn; polyketide aglycone; transgenic plant; insect resistance;
KW macrolide; insecticidal; ds.
XX
XX Saccharopolyspora spinosa.
OS
XX
XX DE1957268-A1.
PN
XX 08-MAR-2001.
PD
XX
XX 29-NOV-1999; 99DE-1057268.
PF

```

```

XX 27-AUG-1999; 99DE-1040596.
XX (FARB ) BAYER AG.
XX Eberz G, Moehrle V, Froede R, Velten R, Salas JA;
XX WPI; 2001-267102/28.
XX
XX New nucleic acid encoding enzymes for spinosyn biosynthesis, useful for
XX recombinant production of insecticidal spinosyns and their derivatives
XX
XX Claim 7; Page 58-74; 354pp; German.
XX
XX This invention describes a novel method nucleic acid (I) and its encoded
XX polypeptide (II) containing at least one region that encodes an enzymatic
XX activity involved in biosynthesis of spinosyns. (I) are used (i) to
XX identify, inactivate or modulate genes involved in the biosynthesis of
XX (II); (ii) to generate a library of polyketide synthases; (iii) for
XX adding forosamine or trimethylrhannose to a spinosyn or polyketide
XX aglycone; and (iv) for recombinant production of the corresponding
XX enzymes, which are used for production of (II), their precursors or
XX derivatives, including production of transgenic plants that express (II)
XX and thus have increased resistance to insects. (I) are also useful as
XX markers for sequencing of the Saccharopolyspora spinosa genome. (II) are
XX macrolides with insecticidal, but not antibacterial, activity, and can
XX also be used to raise specific antibodies, useful for identifying
XX expression clones in a gene bank. Cells transformed with (I) may produce
XX (II) at significantly increased levels or produce new derivatives of
XX (II). This sequence represents a genomic DNA fragment of the S. spinosa
XX genome which contains the coding regions for proteins involved in
XX forosamine and trimethylrhannose biosynthesis.
XX
XX Sequence 45624 BP; 7933 A; 14913 C; 14940 G; 7838 T; 0 other;

Query Match 2.8%; Score 17; DB 22; Length 45624;
Best Local Similarity 100.0%; Pred. No. 84;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 329 gcgcacctcctggagatcg 345
Db 6789 gcgcacctcctggagatcg 6805

RESULT 45
AAF88312/C
XX AAF88312 standard; DNA; 50000 BP.
XX
XX AAF88312;
XX
XX 28-AUG-2001 (first entry)
XX
XX S. spinosa DNA fragment SEQ ID 1.
XX
XX Forosamine; trimethylrhannose; polyketide synthase; biosynthesis;
XX spinosyn; polyketide aglycone; transgenic plant; insect resistance;
XX macrolide; insecticidal; ds.
XX
XX Saccharopolyspora spinosa.
XX
XX DE19957268-A1.
XX
XX 08-MAR-2001.
XX
XX 29-NOV-1999; 99DE-1057268.
XX
XX 27-AUG-1999; 99DE-1040596.
XX (FARB ) BAYER AG.
XX
XX Eberz G, Moehrle V, Froede R, Velten R, Salas JA;

```

```

XX WPI; 2001-267102/28.
XX
XX New nucleic acid encoding enzymes for spinosyn biosynthesis, useful for
XX recombinant production of insecticidal spinosyns and their derivatives
XX
XX Claim 7; Page 14-31; 354pp; German.
XX
XX This invention describes a novel method nucleic acid (I) and its encoded
XX polypeptide (II) containing at least one region that encodes an enzymatic
XX activity involved in biosynthesis of spinosyns. (I) are used (i) to
XX identify, inactivate or modulate genes involved in the biosynthesis of
XX (II); (ii) to generate a library of polyketide synthases; (iii) for
XX adding forosamine or trimethylrhannose to a spinosyn or polyketide
XX aglycone; and (iv) for recombinant production of the corresponding
XX enzymes, which are used for production of (II), their precursors or
XX derivatives, including production of transgenic plants that express (II)
XX and thus have increased resistance to insects. (I) are also useful as
XX markers for sequencing of the Saccharopolyspora spinosa genome. (II) are
XX macrolides with insecticidal, but not antibacterial, activity, and can
XX also be used to raise specific antibodies, useful for identifying
XX expression clones in a gene bank. Cells transformed with (I) may produce
XX (II) at significantly increased levels or produce new derivatives of
XX (II). This sequence represents a genomic DNA fragment of the S. spinosa
XX genome which contains the coding regions for proteins involved in
XX forosamine, trimethylrhannose and polyketide synthase biosynthesis.
XX
XX Sequence 50000 BP; 8484 A; 16290 C; 16535 G; 8691 T; 0 other;

Query Match 2.8%; Score 17; DB 22; Length 50000;
Best Local Similarity 100.0%; Pred. No. 84;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 329 gcgcacctcctggagatcg 345
Db 38836 GCGCATCTCGAGATCG 38820

Search completed: April 30, 2002, 10:51:14
Job time: 10743 sec

```



-----

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 29, 2002, 19:49:59 ; Search time 888.17 Seconds  
(without alignments)  
927.884 Million cell updates/sec

Title: US-09-248-178-56  
Perfect score: 480  
Sequence: 1 ggctttgaagcattttgtc.....gcaacaaaaaaaaaaaaa 480

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_032802.\*

1:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1980.DAT.*
2:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1981.DAT.*
3:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1982.DAT.*
4:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1983.DAT.*
5:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1984.DAT.*
6:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1985.DAT.*
7:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1986.DAT.*
8:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1987.DAT.*
9:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1988.DAT.*
10:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1989.DAT.*
11:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1990.DAT.*
12:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1991.DAT.*
13:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1992.DAT.*
14:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1993.DAT.*
15:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1994.DAT.*
16:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1995.DAT.*
17:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1996.DAT.*
18:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1997.DAT.*
19:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1998.DAT.*
20:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1999.DAT.*
21:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT.*
22:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT.*
23:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA2001B.DAT.*
24:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result NO.	Score	Query Match	Length	ID	Description
1	480	100.0	480	20	AA84202
2	480	100.0	480	21	AA84202
3	480	100.0	489	20	AA84202
4	480	100.0	489	20	AA84202
5	480	100.0	578	21	AA84202
6	480	100.0	578	22	AA84202
7	478.4	99.7	553	19	AA84202
8	478.4	99.7	553	19	AA84202
9	476.4	99.2	497	20	AA84202

10	174.8	58.7	544	21	AA84202	Human breast speci
11	469.4	57.8	471	21	AA84202	Membrane-bound pro
12	469.4	57.8	471	22	AA84202	Human PRO1160 (UNQ
13	469.4	57.8	471	22	AA84202	Human angioogenesis
14	466.6	57.6	488	22	AA84202	Human breast cance
15	465.2	56.9	482	19	AA84202	BS106 polynucleoti
16	465.2	56.9	482	20	AA84202	BS106 consensus po
17	464.4	56.8	472	22	AA84202	Human BS106 gene-s
18	464.4	56.8	472	22	AA84202	Human BS106 gene-s
19	450	53.8	471	21	AA84202	Breast cancer prot
20	450	53.8	471	22	AA84202	Nucleotide sequenc
21	307	64.0	501	23	AA84202	BS106 polynucleoti
22	304.4	63.4	308	19	AA84202	BS106 polynucleoti
23	304.4	63.4	308	20	AA84202	Human BS106 gene-s
24	304.4	63.4	308	22	AA84202	Human BS106 gene-s
25	288.4	60.1	292	22	AA84202	Breast cancer prot
26	273	56.9	273	21	AA84202	cDNA encoding SRT
27	259.2	54.0	264	22	AA84202	cDNA encoding SRT
28	259.2	54.0	266	22	AA84202	BS106 polynucleoti
29	250.8	52.2	291	22	AA84202	BS106 polynucleoti
30	206.4	43.0	229	19	AA84202	Human BS106 gene-s
31	206.4	43.0	229	20	AA84202	Human BS106 gene-s
32	196	40.8	197	19	AA84202	DNA encoding novel
33	196	40.8	197	20	AA84202	BS106 clone 16288
34	196	40.8	197	22	AA84202	BS106 clone 120981
35	188.4	35.2	201	22	AA84202	Human BS106 gene-s
36	96	20.0	660	23	AA84202	Human BS106 gene-s
C 37	70.2	14.6	397	20	AA84202	DNA encoding novel
C 38	65.4	13.6	477	21	AA84202	Spinocerebellar at
C 39	65	13.5	234	16	AA84202	Human SCA7 genomic
C 40	64.8	13.5	1326	23	AA84202	Spinocerebellar at
C 41	64.8	13.5	3376	16	AA84202	DNA encoding novel
C 42	64.6	13.5	203	19	AA84202	AF-9 cDNA. Homo s
C 43	64.6	13.5	203	203	AA84202	Glutamine rich reg
C 44	64.6	13.5	1037	21	AA84202	SCA2 gene CAG repe
C 45	64.6	13.5	1472	21	AA84202	Exons E, C and A o

#### ALIGNMENTS

RESULT 1  
AA84202  
ID AA84202 standard; cDNA; 480 BP.  
XX  
AC AA84202:  
XX  
DI 08-SEP-1999 (first entry)  
XX  
DE DNA encoding human breast tumour protein immunogenic fragment.  
XX  
DE Breast tumour protein; immunogenic fragment; vaccine; detection;  
XX  
KW Breast cancer development; therapy; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO9933869-A2.  
XX  
PD 08-JUL-1999.  
XX  
PF 22-DEC-1998; 98WO-US27416.  
XX  
PR 17-JUL-1998; 98US-0118627.  
PR 24-DEC-1997; 97US-0598253.  
PR 24-DEC-1997; 97US-0598255.  
PR 17-JUL-1998; 98US-0118554.  
XX  
XX (CORI-) CORIXA CORP.  
PI Reed SG, Xu J.  
XX  
DR WPI; 1999-405486/34.  
XX

PT New breast tumour protein genes used, in vaccines for immunotherapy;  
PT or for diagnosis of breast cancer  
PS Claim 3; Page 58; 70pp; English.  
XX  
XX  
XX This sequence encodes a human breast tumour protein immunogenic fragment  
CC of the invention. The polypeptides or nucleic acids encoding them are  
CC useful in vaccines and pharmaceutical compositions for manufacture of  
CC medicaments for inhibiting the development of breast cancer in a patient.  
CC They can also be used to treat breast cancer. Antibodies against these  
CC polypeptides can be used to detect and monitor progression of breast  
CC cancer in patients. Primers and probes derived from the polynucleotides  
CC encoding the breast proteins are useful for detection of breast cancer.  
CC Peripheral blood cells from a patient incubated in the presence of at  
CC least one polypeptide, such that T cells proliferate, are useful in  
CC manufacture of a medicament for treating breast cancer in a patient.  
CC Antigen presenting cells incubated in the presence of at least one  
CC polypeptide are also useful for treating breast cancer.  
XX  
SQ Sequence 480 BP; 121 A; 132 C; 87 G; 140 T; 0 other;

	Query Match	100.0%;	Score	480;	DB	20;	Length	480;		
	Best Local Similarity	100.0%;	Pred. No.	1.9e-109;						
	Matches	480;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
Qy	1	ggctttgaagcatttttgtgtgctccctgatcttcaggtcaccacacatgaagtcttta	60							
Db	1	ggctttgaagcatttttgtgtgctccctgatcttcaggtcaccacacatgaagtcttta	60							
Qy	61	gcagtcctgtactcttgggaagttccatctttcttggtctctgccagaataccgcacaaca	120							
Db	61	gcagtcctgtactcttgggaagttccatctttcttggtctctgccagaataccgcacaaca	120							
Qy	121	gctgtccagctgcacagctatccagctactggtcctctgtatgatgaagccccgtatgct	180							
Db	121	gctgtccagctgcacagctatccagctactggtcctctgtatgatgaagccccgtatgct	180							
Qy	181	gaaccactgctgtgcacaacctgcagcaactgctgcctctcacactgcacaaccagct	240							
Db	181	gaaccactgctgtgcacaacctgcagcaactgctgcctctcacactgcacaaccagct	240							
Qy	241	gcttctaacactgctgttaagaacatccagtttacccaataatgggttggggatctcccg	300							
Db	241	gcttctaacactgctgttaagaacatccagtttacccaataatgggttggggatctcccg	300							
Qy	301	aatggtcagatggtgccctcgatgcggaatacagcttgagttcttcgtcaattggtcacaaact	360							
Db	301	aatggtcagatggtgccctcgatgcggaatacagcttgagttcttcgtcaattggtcacaaact	360							
Qy	361	attcatgcttccctgattttcatccaaactacttaccttgcctcagatataccccctttatct	420							
Db	361	attcatgcttccctgattttcatccaaactacttaccttgcctcagatataccccctttatct	420							
Qy	421	ctaatacagttattttcttccaaaataaaaaataactatgcagcaacaaaataaaaaaaa	480							
Db	421	ctaatacagttattttcttccaaaataaaaaataactatgcagcaacaaaataaaaaaaa	480							

RESULT	2
AAC79431	
ID	AAC79431 standard; cDNA; 480 BP.
,XX	
.AC	AAC79431;
.XX	
.XX	07-FEB-2001 (first entry)
.XX	
.XX	
.XX	
.XX	cDNA sequence of human breast tumour clone 1015F8.
.XX	
KW	Human; breast tumour antigen; cytostatic; immunotherapy;
KW	breast cancer; vaccine; ss.
.XX	
.XX	
OS	Homo sapiens.

XX	WO2000061756-A2.
PN	
XX	
XX	19-GCT-2000.
PD	
XX	
XX	10-APR-2000; 20JUNJ-US09688.
XX	
XX	05-APR-1999; 99US-0288950.
PR	02-JUL-1999; 99US-0346327.
XX	
XX	(CORI-) CORIXA CORP.
PA	
XX	
XX	Reed SG, Xu J, Dillon DC;
PI	
XX	WPI; 2000-638529/61.
XX	
XX	A novel isolated polypeptide comprising an immunogenic portion of a
PI	breast cancer protein useful in the detection and treatment of breast
PI	cancer -
PI	
XX	
XX	Claim 4; Page 74; 95pp; English.
XX	
XX	The present sequence was isolated from a breast tumour cDNA library. It
CC	is provided in a specification relating to compounds for immunotherapy
CC	and diagnosis of breast cancer. Breast tumour antigens and the
CC	polynucleotides that encode them may be used in the production of a
CC	pharmaceutical composition to be used in the treatment of breast cancer.
CC	Proliferated T cells and incubated antigen presenting cells are also
CC	required. The polypeptides and polynucleotides may also be used to
CC	produce a vaccine.
XX	
XX	Sequence 480 Br; 121 A; 132 C; 87 G; 140 T; 0 other;

Query Match.	100.0%;	Score	480;	DB	21;	Length	480;		
Best Local Similarity	100.0%;	Pred. NO.	1.9e-105;						
Matches	480;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;

  

Qy	1	ggcttgaacatttttctgtctgtccctgatcttcaggtccacacacatgaagtcttta	60
Db	1	ggcttgaagcatttttctgtctgtccctgatcttcaggtccacacacatgaagtcttta	60
Qy	61	gcagtcctgttactcttgggagttccatcttttctgtctgtcccgagatcccgacaaca	120
Db	61	gcagtcctgttactcttgggagttccatcttttctgtctgtcccgagatcccgacaaca	120
Qy	121	gctgtccagctgacacagatccagctactgtgtctgtctgatgatgaagccctgatgct	180
Db	121	gctgtccagctgacacagatccagctactgtgtctgtctgatgatgaagccctgatgct	180
Qy	181	gaagccactgttctgaaacacactgcgaccactgctgtctctaccactgcaaccacgct	240
Db	181	gaagccactgttctgaaacacactgcgaccactgctgtctctaccactgcaaccacgct	240
Qy	241	gcttctaccctgtctgttaagacattccagtttaccacaaatgggttgggattctcccg	300
Db	241	gcttctaccctgtctgttaagacattccagtttaccacaaatgggttgggattctcccg	300
Qy	301	aatggtagatgtgtccctggagatggaaatcagcttgagcttctgtcaatttggtcacaaact	360
Db	301	aatggtagatgtgtccctggagatggaaatcagcttgagcttctgtcaatttggtcacaaact	360
Qy	361	attcatgcttctgtgatttcatccaaacttacttacttgcctagagataccccctttatct	420
Db	361	attcatgcttctgtgatttcatccaaacttacttacttgcctagagataccccctttatct	420
Qy	421	ctaatcaggtttattttcttcaataaaaaataactatgatgacacaaaaaataaaaaa	480
Db	421	ctaatcaggtttattttcttcaataaaaaataactatgatgacacaaaaaataaaaaa	480

RESULT 3

AAX00704



PK 30-MAY-1997; 97US-0050937.  
PR 21-MAR-1997; 97US-0041276.  
PR 21-MAR-1997; 97US-0041277.  
PR 21-MAR-1997; 97US-0041281.  
PR 21-MAR-1997; 97US-0042344.  
PR 30-MAY-1997; 97US-0048069.  
PR 30-MAY-1997; 97US-0048094.  
PR 30-MAY-1997; 97US-0048095.  
PR 30-MAY-1997; 97US-0048096.  
PR 30-MAY-1997; 97US-0048099.  
PR 30-MAY-1997; 97US-0048131.  
PR 30-MAY-1997; 97US-0048135.  
PR 30-MAY-1997; 97US-0048154.  
PR 30-MAY-1997; 97US-0048160.  
PR 30-MAY-1997; 97US-0048186.  
PR 30-MAY-1997; 97US-0048187.  
PR 30-MAY-1997; 97US-0048188.  
PR 30-MAY-1997; 97US-0048350.  
PR 30-MAY-1997; 97US-0048351.  
PR 30-MAY-1997; 97US-0048352.  
PR 30-MAY-1997; 97US-0048355.  
PR 05-AUG-1997; 97US-0054804.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Brewer LA, Duan R, Ebner R, Ferrie AM, Florence KA;  
PI Greene JM, Hu JS, Lafleur DW, Moore PA, Ni J, Olsen HS;  
PI Rosen CA, Ruben SM, Shi Y, Young P;  
XX  
XX WPI; 1999-070066/06.  
DR P-PSDB; AAW67840.  
XX  
XX New isolated human genes and the secreted polypeptides they encode -  
PI useful for diagnosis and treatment of e.g. cancers, neurological  
PI disorders, immune diseases, inflammation or blood disorders  
XX  
XX Claim 1; Page 198; 385pp; English.  
XX  
XX This sequence represents a nucleic acid molecule which encodes a  
CC secreted human protein. The gene number, and the clone it is derived  
CC from, are detailed in the descriptor line. The gene can be used to  
CC generate fusion proteins by linking to the gene to a human immunoglobulin  
CC Fc portion (e.g. AAX00602) for increasing the stability of the fused  
CC protein as compared to the human protein only.  
CC The invention relates to 87 novel genes and their fragments (nucleic  
CC acid sequences: AAX00611-X00724; amino acid sequences AAW67807-W68004)  
CC which are useful for preventing, treating or ameliorating medical  
CC conditions e.g. by protein or gene therapy. Also, pathological  
CC conditions can be diagnosed by determining the amount of the new  
CC polypeptides in a sample or by determining the presence of mutations in  
CC the new polynucleotides. Specific uses are described for each of the 87  
CC polynucleotides, based on which tissues they are most highly expressed in  
CC (see AAX00611 for described uses).  
XX  
XX Sequence 489 BP; 125 A; 134 C; 87 G; 143 T; 0 other;

Query Match 100.0%; Score 480; DB 20; Length 489;  
Best Local Similarity 100.0%; Pred. No. 1.9e-109;  
Matches 480; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ggccttgagcattttgtgtgtccctgattcttcagggtccaccacatgaagtcttta 60  
DB 7 ggccttgagcattttgtgtgtccctgattcttcagggtccaccacatgaagtcttta 66  
QY 61 ggaagtcctggtactcttgggagttccatcttctgtctgtccagaatccgacaca 120  
DB 67 ggaagtcctggtactcttgggagttccatcttctgtctgtccagaatccgacaca 126  
QY 121 gctgtccagctgacacatccagactactggtctgtgatgatgaagccctgatgct 180  
DB 127 gctgtccagctgacacatccagactactggtctgtgatgatgaagccctgatgct 186

QY 181 gaaccactgtgtgcgaaccactgcgaccactgtgtctctaccactgcgaaccacgct 240  
DB 187 gaaccactgtgtgcgaaccactgcgaccactgtgtctctaccactgcgaaccacgct 246  
QY 241 gcttctaccactgtcgtlaaagacattccagttttaccacaaatgggttggggtatctcccg 300  
DB 247 gcttctaccactgtcgtlaaagacattccagttttaccacaaatgggttggggtatctcccg 306  
QY 301 aalggtagagtgtgctccctgagatggaatcagcttgagcttcttgcgaattggtcacaaact 360  
DB 307 aalggtagagtgtgctccctgagatggaatcagcttgagcttcttgcgaattggtcacaaact 366  
QY 361 attcargcttctgtgatttccatccactacttaccactgttcacatccctttatct 420  
DB 367 attcargcttctgtgatttccatccactacttaccactgttcacatccctttatct 426  
QY 421 ctaatcagtttattttcttccaaataaaaaataaactatgagcaacaaaaa 480  
DB 427 ctaatcagtttattttcttccaaataaaaaataaactatgagcaacaaaaa 486  
RESULT 5  
AAC79470  
ID AAC79470 standard; cDNA; 578 BP.  
XX  
XX AC AAC79470;  
XX  
XX AC AAC79470;  
DT 07-FEB-2001 (first entry)  
XX  
XX cDNA sequence of human breast tumour clone B511S.  
XX  
XX Human; breast tumour antigen; cytostatic; immunotherapy;  
KW breast cancer; vaccine; ss.  
XX  
XX OS Homo sapiens.  
XX  
XX WO200061756-A2.  
XX  
XX 19-OCT-2000.  
PD  
XX  
PF 10-APR-2000; 2000WO-US09688.  
XX  
XX 09-APR-1999; 99US-0288950.  
PR 02-JUL-1999; 99US-0346327.  
XX  
XX (CORI-) CORIXA CORP.  
PA  
XX  
XX Reed SG, Xu J, Dillon DC;  
XX  
XX WPI; 2000-638568/61.  
DR P-PSDB; AAS28525.  
XX  
XX A novel isolated polypeptide comprising an immunogenic portion of a  
PI breast cancer protein useful in the detection and treatment of breast  
PI cancer -  
XX  
XX Claim 4; Page 88; 95pp; English.  
XX  
XX The present sequence was isolated from a breast tumour cDNA library. It  
CC is provided in a specification relating to compounds for immunotherapy  
CC and diagnosis of breast cancer. Breast tumour antigens and the  
CC polynucleotides that encode them may be used in the production of a  
CC pharmaceutical composition to be used in the treatment of breast cancer.  
CC Proliferated I cells and incubated antigen presenting cells are also  
CC required. The polypeptides and polynucleotides may also be used to  
CC produce a vaccine.  
XX  
XX Sequence 578 BP; 206 A; 137 C; 88 G; 147 T; 0 other;

Query Match 100.0%; Score 480; DB 21; Length 578;  
Best Local Similarity 100.0%; Pred. No. 2e-109;  
Matches 480; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	ggctttgaagcatttttctgtctgtccctgatctctcaggtcaccacccatgaagttctta	60
Db	15	ggctttgaagcatttttctgtctgtccctgatctctcaggtcaccacccatgaagttctta	74
QY	61	gcagtcctggtactcttgagagttcccatctttcttggtctctgccagaaatccgcagaca	120
Db	75	gcagtcctggtactcttgagagttcccatctttcttggtctctgccagaaatccgcagaca	134
QY	121	gctgtctcagctgacacgtatccagctactggtctctgtgatgatgaagccctctgatgt	180
Db	135	gctgtctcagctgacacgtatccagctactggtctctgtgatgatgaagccctctgatgt	194
QY	181	gaacacactgctactgaacacactgcgacacactgctctctcactcaccactcaaccacgct	240
Db	195	gaacacactgctactgaacacactgcgacacactgctctctcactcaccactcaaccacgct	254
QY	241	gctctcaccactgctctgaagacattccagttttaccacaaatgggttggtggtatcccg	300
Db	255	gctctcaccactgctctgaagacattccagttttaccacaaatgggttggtggtatcccg	314
QY	301	aatggtagagtggtctccctgagatggaaatcagcttgagctctctgcgaatggtgcacaact	360
Db	315	aatggtagagtggtctccctgagatggaaatcagcttgagctctctgcgaatggtgcacaact	374
QY	361	attcatgctctctgtgatttccatccaaactacttactcttgcctacgataccccctttatct	420
Db	375	attcatgctctctgtgatttccatccaaactacttactcttgcctacgataccccctttatct	434
QY	421	ctaatacagtt	480
Db	435	ctaatacagtt	494
RESULT 7			
AAV31993			
ID	AAV31993	standard; DNA: 553 BP.	
XX	XX		
AC	AAV31993;		
DT	25-SEP-1998	(first entry)	
XX	XX		
DE	BS106	polynucleotide clone 1662885 (ii).	
XX	XX		
KW	ss; human; BS106;	breast tissue gene; breast cancer; detection marker.	
XX	XX		
OS	Homo sapiens.		
QY	1	ggctttgaagcatttttctgtctgtccctgatctctcaggtcaccacccatgaagttctta	60
Db	15	ggctttgaagcatttttctgtctgtccctgatctctcaggtcaccacccatgaagttctta	74
QY	61	gcagtcctggtactcttgagagttcccatctttcttggtctctgccagaaatccgcagaca	120
Db	75	gcagtcctggtactcttgagagttcccatctttcttggtctctgccagaaatccgcagaca	134
QY	121	gctgtctcagctgacacgtatccagctactggtctctgtgatgatgaagccctctgatgt	180
Db	135	gctgtctcagctgacacgtatccagctactggtctctgtgatgatgaagccctctgatgt	194
QY	181	gaacacactgctactgaacacactgcgacacactgctctctcactcaccactcaaccacgct	240
Db	195	gaacacactgctactgaacacactgcgacacactgctctctcactcaccactcaaccacgct	254
QY	241	gctctcaccactgctctgaagacattccagttttaccacaaatgggttggtggtatcccg	300
Db	255	gctctcaccactgctctgaagacattccagttttaccacaaatgggttggtggtatcccg	314
QY	301	aatggtagagtggtctccctgagatggaaatcagcttgagctctctgcgaatggtgcacaact	360
Db	315	aatggtagagtggtctccctgagatggaaatcagcttgagctctctgcgaatggtgcacaact	374
QY	361	attcatgctctctgtgatttccatccaaactacttactcttgcctacgataccccctttatct	420
Db	375	attcatgctctctgtgatttccatccaaactacttactcttgcctacgataccccctttatct	434
QY	421	ctaatacagtt	480
Db	435	ctaatacagtt	494
AAI67224			
ID	AAI67224	standard; cDNA; 578 BP.	
XX	XX		
AC	AAI67224;		
DT	11-FEB-2002	(first entry)	
XX	XX		
DE	BS11S	cDNA sequence.	
XX	XX		
KW	Genetic subtraction; DNA microarray analysis;	polymerase chain reaction;	
KW	cancer; BS11S; ss.		
XX	XX		
OS	Homo sapiens.		
XX	XX		
FH	Key	Location/Qualifiers	
FT	CDS	63..335	
FT		/*tag= a	
XX	XX		
PN	WO200175171-A2.		
PD	11-OCT-2001.		
XX	XX		
PF	02-APR-2001; 2001WO-US10631.		
XX	XX		
PR	03-APR-2000; 2000US-194241P.		
PR	20-JUL-2000; 2000US-219862P.		
PR	27-JUL-2000; 2000US-221300P.		
PR	18-DEC-2000; 2000US-256592P.		
XX	XX		
PA	(CORI-) CORIXA CORP.		
XX	XX		
PI	Houghton RL, Dillon DC, Molesh DA, Xu J, Zehentner B, Persing DH;		
XX	XX		
XX	WPI; 2001-626449/72.		
DR	P-PSDB; AAG65988.		
XX	XX		
PT	Identifying tissue (tumour)-specific polynucleotides overexpressed in		
PT	tissue of interest as compared to control tissue, for detecting cancer		

PI cells in patient, comprises DNA microarray analysis or quantitative  
PI polymerase chain reaction -  
XX  
PS Claim 4; Page 116; 127pp; English.  
XX  
CC The invention relates to identifying tissue-specific polynucleotides (P)  
CC that involves performing a genetic subtraction to identify pool of (P)  
CC from tissue of interest (TI), performing DNA microarray analysis to  
CC identify first subset of polynucleotides (SPI) at least 2-fold over  
CC expressed in TI, and performing quantitative polymerase chain reaction  
CC (PCR) analysis on SPI to identify second subset of (P). The method is  
CC useful for determining the presence or absence of a cancer cell in a  
CC patient, monitoring the progression of cancer in a patient using a  
CC biological sample such as blood, serum, lymph nodes, bone marrow, sputum,  
CC urine or a tumour biopsy sample. The methods are useful for determining  
CC the presence or absence of or monitoring progression of prostate, breast,  
CC colon, ovarian, lung, head and neck, lymphoma, leukemia, melanoma, liver,  
CC gastric, kidney, bladder, pancreatic or endometrial cancer. The present  
CC sequence represents BS11S cDNA.

SQ Sequence 578 BP; 206 A; 137 C; 88 G; 147 T; 0 other;

Query Match 100.0%; Score 480; DB 22; Length 578;  
Best Local Similarity 100.0%; Pred. No. 2e-109;  
Matches 480; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggctttgaagcatttttctgtctgtccctgatctctcaggtcaccacccatgaagttctta 60  
Db 15 ggctttgaagcatttttctgtctgtccctgatctctcaggtcaccacccatgaagttctta 74

QY 61 gcagtcctggtactcttgagagttcccatctttcttggtctctgccagaaatccgcagaca 120  
Db 75 gcagtcctggtactcttgagagttcccatctttcttggtctctgccagaaatccgcagaca 134

QY 121 gctgtctcagctgacacgtatccagctactggtctctgtgatgatgaagccctctgatgt 180  
Db 135 gctgtctcagctgacacgtatccagctactggtctctgtgatgatgaagccctctgatgt 194

QY 181 gaacacactgctactgaacacactgcgacacactgctctctcactcaccactcaaccacgct 240  
Db 195 gaacacactgctactgaacacactgcgacacactgctctctcactcaccactcaaccacgct 254

QY 241 gctctcaccactgctctgaagacattccagttttaccacaaatgggttggtggtatcccg 300  
Db 255 gctctcaccactgctctgaagacattccagttttaccacaaatgggttggtggtatcccg 314

QY 301 aatggtagagtggtctccctgagatggaaatcagcttgagctctctgcgaatggtgcacaact 360  
Db 315 aatggtagagtggtctccctgagatggaaatcagcttgagctctctgcgaatggtgcacaact 374

QY 361 attcatgctctctgtgatttccatccaaactacttactcttgcctacgataccccctttatct 420  
Db 375 attcatgctctctgtgatttccatccaaactacttactcttgcctacgataccccctttatct 434

QY 421 ctaatacagtt 480  
Db 435 ctaatacagtt 494

RESULT 7  
AAV31993  
ID AAV31993 standard; DNA: 553 BP.  
XX  
AC AAV31993;  
DT 25-SEP-1998 (first entry)  
XX  
DE BS106 polynucleotide clone 1662885 (ii).  
XX  
KW ss; human; BS106; breast tissue gene; breast cancer; detection marker.  
XX  
OS Homo sapiens.

[illegible]

Db 390 attcaggtcccggtattctcaccacactcacttacttgccctacgataccccctttatct 449  
Qy 421 ctaatcagttattttcttccaaataaaaaataaactatgagcaacaaaaaataaaaaa 480  
Db 450 ctaatcagttattttcttccaaataaaaaataaactatgagcaacaaaaaataaaaaa 509

RESULT 9  
ID AAZ40783  
XX AAZ40783 standard; DNA; 497 BP.  
AC AAZ40783;  
XX  
DT 18-JAN-2000 (first entry)  
XX  
DE Secreted protein EST coding sequence 108-002-5-0-F3-FL.  
XX  
KW Chromosome mapping; fingerprint identification technique;  
KW hyperlipidemia; human; hereditary disease; diagnosis; cancer;  
KW autoimmune disease; rheumatic disease; embryogenic disorder; myopathy;  
KW renal injury; amino aciduria; hypoglycaemia; male rat infertility;  
KW hypertension; ss.  
XX  
OS Homo sapiens.  
XX  
PN W09940189-A2.  
XX  
PD 12-AUG-1999.  
XX  
PF 09-FEB-1999; 99WO-IB00282.  
XX  
PR 09-FEB-1998; 98US-0074121.  
PR 13-APR-1998; 98US-0081563.  
PR 10-AUG-1998; 98US-0096116.  
PR 04-SEP-1998; 98US-0099273.  
XX  
PA (GEST ) GENSET.  
XX  
PI Bougueleret L, Duclert A, Dumas Milne Edwards J;  
XX  
DR WPI; 1999-600966/51.  
DR P-PSDB; AAY59655.  
XX  
PT Extended cDNAs useful for expressing secreted proteins and to obtain  
PT specific antibodies -  
XX  
PS Claim 1; Page 150; 244pp; English.  
XX  
CC This sequence encodes a human secreted protein of the invention. The  
CC extended cDNAs (or genomic DNAs obtainable from them) may be used to  
CC prepare PCR primers and probes. These are useful for forensic matching or  
CC positive identification by DNA sequencing. They may also be used in  
CC alternative fingerprint identification techniques. Antibodies against the  
CC proteins encoded by the extended cDNAs are useful in identification of  
CC tissue types or cell species, as well as identifying tissue specific  
CC soluble proteins. The sequences can be used for chromosome mapping and  
CC identification of genes associated with hereditary diseases or drug  
CC response. signal sequences from the cDNAs can be used in construction of  
CC secretion vectors. Other sequences derived from the extended cDNAs can be  
CC used to clone upstream genomic DNA sequences including promoters. This is  
CC in turn useful for identifying proteins that interact with promoter  
CC sequences. Some of the proteins may be useful in diagnosing and treating  
CC several disorders including, but not limited to: cancer, hyperlipidaemia,  
CC cardiovascular and neurodegenerative disorders, autoimmune diseases, and  
CC rheumatic diseases, embryogenic disorders, hypertension, renal injury,  
CC amino acidurias, hypoglycaemia, male rat infertility and myopathies.  
XX  
SQ Sequence 497 BP; 122 A; 138 C; 90 G; 147 T; 0 other;

Query Match 99.2%; Score 476.4; DB 20; Length 497;

Best local Similarity 59.8%; Pred. No. 1.5e-108;  
Matches 477; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ggctttgagcaatttttctgtctgtccctgatcttccaggtccaccaccatgaagtcttta 60  
Db 20 ggctttgagcaatttttctgtctgtccctgatcttccaggtccaccaccatgaagtcttta 79  
Qy 61 gcagtcctcggtaactctctggagatttcccatcttcttggtctctgcccagaatccgacaaca 120  
Db 80 gcagtcctcggtaactctctggagatttcccatcttcttggtctctgcccagaatccgacaaca 139  
Qy 121 gctgtccagctgacgcgtatccagctactgctctctgtctgtgatgaaagccctctgatgt 180  
Db 140 gctgtccagctgacgcgtatccagctactgctctctgtctgtgatgaaagccctctgatgt 199  
Qy 161 gaacccaactgctgtgcaacccaactgcgaccactgcgaccactgcgtctctaccactgcaacccaactgct 240  
Db 200 gaacccaactgctgtgcaacccaactgcgaccactgcgtctctaccactgcaacccaactgct 259  
Qy 241 gctttaccactgcgtcgttaaaagacattccagttttaccacaaatgggttgaggatctcccg 300  
Db 260 gctttaccactgcgtcgttaaaagacattccagttttaccacaaatgggttgaggatctcccg 319  
Qy 301 aaaggtagagtggtgctccctgagatggaatcagcttgagcttctctgcaattggtcacaaact 360  
Db 320 aaaggtagagtggtgctccctgagatggaatcagcttgagcttctctgcaattggtcacaaact 379  
Qy 361 attcatgcttctctgttatttctccaaactacttaacttgcctgctacgataatcccttttatct 420  
Db 380 attcatgcttctctgttatttctccaaactacttaacttgcctgctacgataatcccttttatct 439  
Qy 421 cttaatcagttattttcttccaaataaaaaataaactatgagcaacaaaaaataaaaaa 478  
Db 440 cttaatcagttattttcttccaaataaaaaataaactatgagcaacaaaaaataaaaaa 497

RESULT 10  
AAZ51765  
ID AAZ51765 standard; DNA; 544 BP.  
XX  
AC AAZ51765;  
XX  
DI 01-JUN-2000 (first entry)  
XX  
DE Human breast specific gene clone Mam002.  
XX  
KW Human; breast specific gene; breast specific marker; BSG; diagnosis;  
KW breast cancer; therapy; Mam002; ss.  
XX  
OS Homo sapiens.  
XX  
PN W0200008210-A1.  
XX  
PD 17-FEB-2000.  
XX  
PF 22-JUL-1999; 99WO-US16811.  
XX  
PR 04-AUG-1998; 98US-0095232.  
XX  
PA (DIAD-) DIADEXUS LLC.  
XX  
PI Sun Y, Recipon H, Cafferkey R;  
XX  
DR WPI; 2000-205737/18.  
XX  
KW Detecting, diagnosing, monitoring, staging, prognosticating, imaging  
PI and treating breast cancer using protein product of breast specific  
PI genes -  
XX  
PS Claim 7; Page 38; 45pp; English.  
XX  
CC This sequence represents a breast specific gene (BSG) clone. The BSG  
CC sequences are also referred to as breast specific markers (BSM). The



CC invention relates to methods for diagnosing, monitoring, staging, imaging  
CC and treating breast cancer (BC). The methods comprise measuring the  
CC levels of BSG products in cells, tissues or body fluids of the patient  
CC and comparing the measured levels of BSG, with BSG levels of a normal  
CC human control. An antibody against the BSG sequences can be labelled and  
CC used for imaging BC in a patient. The antibody can be conjugated to a  
CC cytotoxic agent, and used for treating BC in a patient.

XX  
SQ Sequence 544 BP; 142 A; 143 C; 94 G; 154 T; 11 other;

Query Match 98.7%; Score 473.8; DB 21; Length 544;  
Best Local Similarity 99.0%; Pred. No. 6.8e-108;  
Matches 475; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 ggtttgaaacatttttctgtgctccctgactcttcctcaggtccaccaccatgaagttcttta 60  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 61 gcagtcctgtaactttggaggatttccatctttctggtctctgccagaaatccgcacaaca 120  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 96 gcagtcctgtaactttggaggatttccatctttctggtctctgccagaaatccgcacaaca 155  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 121 gctgctccagctgaacagctatccagctactgctctgctgatgatgaagcccttgatgct 180  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 156 gctgctccagctgaacagctatccagctactgctctgctgatgatgaagcccttgatgct 215  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 181 gaaacactgctgctgaacactgcgacactgctgctctcctaccactgcaaccacgct 240  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 216 gaaacactgctgctgaacactgcgacactgctgctctcctaccactgcaaccacgct 275  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 241 gcttctaccactgctgctgaacactgcagctcttaccacaaatgggtgggagatctcccg 300  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 276 gcttctaccactgctgctgaacactgcagctcttaccacaaatgggtgggagatctcccg 335  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 301 aatggtagagtgctccctgagatgaatacagcttgctctcagttgctcaattggtcacaact 360  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 336 aatggtagagtgctccctgagatgaatacagcttgctctcagttgctcaattggtcacaact 395  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 361 attcatgctctctgctgatttccatcccaacttacttactctgctcactgatactccctttatct 420  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 396 attcatgctctctgctgatttccatcccaacttacttactctgctcactgatactccctttatct 455  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 421 ctaatcagttattttcttccaaataaaaaataactatgacgaacaaaaaataaaaaaa 480  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 456 ctaatcagttattttcttccaaataaaaaataactatgacgaacaaaaaataaaaaaa 515  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

RESULT 11

AAZ65099

ID AAZ65099 standard; cDNA; 471 BP.

XX

AC AAZ65099;

XX

XX 05-APR-2000 (first entry)

DT

DE Membrane-bound protein Prol160 encoding cDNA.

XX

XX Membrane-bound polypeptide; PRO polypeptide; LDL receptor; TIE ligand;

KW pharmaceutical; receptor immunoadhesin; gene mapping; ss.

XX

OS Homo sapiens.

XX

XX WU963088-A2.

PN

XX 09-DEC-1999.

PD

XX 02-JUN-1999; 99MO-US12252.

PF

XX 02-JUN-1998; 98US-0087607.

XX

PR 02-JUN-1998; 98US-0087609.

PR

PR 02-JUN-1998; 98US-0087759.

PR

PR 03-JUN-1998; 98US-0087827.

PR 04-JUN-1998; 98US-0088021.  
PR 04-JUN-1998; 98US-0088025.  
PR 04-JUN-1998; 98US-0088028.  
PR 04-JUN-1998; 98US-0088029.  
PR 04-JUN-1998; 98US-0088030.  
PR 04-JUN-1998; 98US-0088033.  
PR 04-JUN-1998; 98US-0088326.  
PR 05-JUN-1998; 98US-0088167.  
PR 05-JUN-1998; 98US-0088202.  
PR 05-JUN-1998; 98US-0088212.  
PR 05-JUN-1998; 98US-0088217.  
PR 05-JUN-1998; 98US-0088655.  
PR 10-JUN-1998; 98US-0088722.  
PR 10-JUN-1998; 98US-0088730.  
PR 10-JUN-1998; 98US-0088734.  
PR 10-JUN-1998; 98US-0088738.  
PR 10-JUN-1998; 98US-0088740.  
PR 10-JUN-1998; 98US-0088741.  
PR 10-JUN-1998; 98US-0088742.  
PR 10-JUN-1998; 98US-0088810.  
PR 10-JUN-1998; 98US-0088811.  
PR 10-JUN-1998; 98US-0088824.  
PR 10-JUN-1998; 98US-0088825.  
PR 10-JUN-1998; 98US-0088826.  
PR 11-JUN-1998; 98US-0088858.  
PR 11-JUN-1998; 98US-0088861.  
PR 11-JUN-1998; 98US-0088863.  
PR 11-JUN-1998; 98US-0088876.  
PR 12-JUN-1998; 98US-0089090.  
PR 12-JUN-1998; 98US-0089105.  
PR 16-JUN-1998; 98US-0089440.  
PR 16-JUN-1998; 98US-0089512.  
PR 16-JUN-1998; 98US-0089514.  
PR 17-JUN-1998; 98US-0089532.  
PR 17-JUN-1998; 98US-0089538.  
PR 17-JUN-1998; 98US-0089598.  
PR 17-JUN-1998; 98US-0089599.  
PR 17-JUN-1998; 98US-0089600.  
PR 17-JUN-1998; 98US-0089653.  
PR 18-JUN-1998; 98US-0089801.  
PR 18-JUN-1998; 98US-0089907.  
PR 18-JUN-1998; 98US-0089908.  
PR 19-JUN-1998; 98US-0089947.  
PR 19-JUN-1998; 98US-0089948.  
PR 19-JUN-1998; 98US-0089952.  
PR 22-JUN-1998; 98US-0090246.  
PR 22-JUN-1998; 98US-0090252.  
PR 22-JUN-1998; 98US-0090254.  
PR 23-JUN-1998; 98US-0090349.  
PR 23-JUN-1998; 98US-0090355.  
PR 24-JUN-1998; 98US-0090429.  
PR 24-JUN-1998; 98US-0090431.  
PR 24-JUN-1998; 98US-0090435.  
PR 24-JUN-1998; 98US-0090444.  
PR 24-JUN-1998; 98US-0090445.  
PR 24-JUN-1998; 98US-0090461.  
PR 24-JUN-1998; 98US-0090535.  
PR 24-JUN-1998; 98US-0090538.  
PR 24-JUN-1998; 98US-0090540.  
PR 24-JUN-1998; 98US-0090557.  
PR 25-JUN-1998; 98US-0090676.  
PR 25-JUN-1998; 98US-0090678.  
PR 25-JUN-1998; 98US-0090688.  
PR 25-JUN-1998; 98US-0090690.  
PR 25-JUN-1998; 98US-0090691.  
PR 25-JUN-1998; 98US-0090694.  
PR 25-JUN-1998; 98US-0090695.  
PR 25-JUN-1998; 98US-0090696.  
PR 26-JUN-1998; 98US-0090862.  
PR 26-JUN-1998; 98US-0090863.  
PR 01-JUL-1998; 98US-0091358.  
PR 01-JUL-1998; 98US-0091360.



```

XX PF 30-MAR-2000; 2000WO-US08439.
XX PR 02-JUN-1999; 99WO-US12252.
XX PR 23-JUN-1999; 99US-0141037.
PR 07-JUL-1999; 99US-0143048.
PR 20-JUL-1999; 99US-0144758.
PR 26-JUL-1999; 99US-0145698.
PR 28-JUL-1999; 99US-0146222.
PR 17-AUG-1999; 99US-0149396.
PR 15-SEP-1999; 99WO-US21090.
PR 15-SEP-1999; 99WO-US21547.
PR 08-OCT-1999; 99US-0158663.
PR 30-NOV-1999; 99WO-US28313.
PR 01-DEC-1999; 99WO-US28301.
PR 16-DEC-1999; 99WO-US30095.
PR 20-DEC-1999; 99WO-US30911.
PR 05-JAN-2000; 2000WO-US00219.
PR 06-JAN-2000; 2000WO-US00376.
PR 11-FEB-2000; 2000WO-US03565.
PR 18-FEB-2000; 2000WO-US04341.
PR 22-FEB-2000; 2000WO-US04414.
PR 24-FEB-2000; 2000WO-US04914.
PR 02-MAR-2000; 2000WO-US05004.
PR 15-MAR-2000; 2000WO-US05941.
PR 20-MAR-2000; 2000WO-US06884.
XX PR 20-MAR-2000; 2000WO-US07377.
PA (GETH ) GENENTECH INC.
XX PI Ashkenazi AJ, Baker KP, Botstein D, Desnoyers L, Eaton DL;
PI Ferrara N, Fong S, Gerber H, Gerritsen ME, Goddard A, Godowski PJ;
PI Grimaldi CJ, Gurney AL, Kijavini LJ, Napier MA, Pan J, Paoni NF;
PI Roy MA, Stewart TA, Tumas D, Watanabe CK, Williams PM, Wood WI;
PI Zhang Z;
XX PI WPI; 2001-032160/04.
DR P-PSDB; AAB65276.
XX PR
XX PR PRO polynucleotides used to produce polypeptides used to target
PT Bioactive molecules such as toxins, radiolabels or antibodies, to
PT specific cells, to cause targeted cell death -
XX PI Claim 2; Fig 281; 935pp; English.
XX PS
XX CC The present invention describes human secreted and transmembrane PRO
CC proteins. The PRO proteins have cytostatic activity. The PRO proteins
CC can be used for targeted delivery of bioactive molecules, such as
CC toxins, radiolabels or antibodies, that cause cell death. PRO nucleotide
CC sequences, and their fragments, can be used as hybridisation probes, in
CC chromosomal and gene mapping, and in the generation of anti-sense RNA
CC and DNA. They may also be used to produce transgenic animals which are
CC used to develop and screen therapeutically useful reagents. The PRO
CC nucleotide and protein sequence can be used for tissue typing and in
CC treating cancer. Anti-PRO antibodies can be used in diagnostic assays.
CC AAF44270 to AAF4470 represent PCR primers and hybridisation probes used
CC in the isolation of human PRO sequences. AAF44087 to AAF44265 and
CC AAB65154 to AAB65300 represent human PRO polynucleotide and protein
CC sequences given in the exemplification of the present invention.
XX PR
XX SQ Sequence 471 BP; 118 A; 131 C; 84 G; 138 T; 0 other;

Query Match 97.8%; Score 469.4; DB 22; Length 471;
Best Local Similarity 99.8%; Pred. No. 7.9e-107;
Matches 470; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 10 gcattttgtctgtccctgtatcttcagggtccaccaccatgaagtcttcagtcctg 69
Db 1 gcatctttgtctgtccctgtatcttcagggtccaccaccatgaagtcttcagtcctg 60
Qy 70 gtactcttggaggtttccattcttctgtctgtccaggaatccgacaacagctgtcca 129

```

```

Db 61 gtactcttggaggtttccattcttctgtctgtccaggaatccgacaacagctgtcca 120
Qy 130 gctgacagctatccagctactggtctctgtgatgatgaagcccttgatgctgaaacct 189
Db 121 gctgacagctatccagctactggtctctgtgatgatgaagcccttgatgctgaaacct 180
Qy 190 gctgctgcaaacactgcgcaccactgctctctaccactgcaaccccgctctctacc 249
Db 181 gctgctgcaaacactgcgcaccactgctctctaccactgcaaccccgctctctacc 240
Qy 250 actgctcgtaaagacattccagttttaccaccaatgggttggggatctcccgaaatgtaga 309
Db 241 actgctcgtaaagacattccagttttaccaccaatgggttggggatctcccgaaatgtaga 300
Qy 310 gctgtccctcgagatgggaatcagctgagctctctgcaattcgttcacacactatcagct 369
Db 301 gctgtccctcgagatgggaatcagctgagctctctgcaattcgttcacacactatcagct 360
Qy 370 tctgtgatttccatcaactcacttacccttgcctacgatacccccttctcttaatacag 429
Db 361 tctgtgatttccatcaactcacttacccttgcctacgatacccccttctcttaatacag 420
Qy 430 ttattttcttccaaataaataactatgagcaacaaaaa 480
Db 421 ttattttcttccaaataaataactatgagcaacaaaaa 471

RESULT 13
AAC97492
ID AAC97492 standard; cDNA; 471 BP.
XX AC AAC97492;
XX DT 26-FEB-2001 (first entry)
XX DE Human angiogenesis-associated protein Proll60 cDNA, SEQ ID NO:159.
XX KW Human; angiogenesis-associated protein; PRO; endothelial cell growth;
KW cardiac hypertrophy; cardiovascular disorder; endothelial disorder;
KW angiogenic disorder; atherosclerosis; osteoporosis; hypertension;
KW myocardial infarction; diabetic retinopathy; rheumatoid arthritis;
KW Crohn's disease; psoriasis; endometriosis; ulcer; wound healing; cancer;
KW Alzheimer's disease; Huntington's disease; stroke; drug screening;
KW gene therapy; transgenic animal; ss.
XX OS Homo sapiens.
XX PN WO2000053753-A2.
XX PD 14-SEP-2000.
XX PF 05-JAN-2000; 2000WO-US00219.
XX PR 06-MAR-1999; 99WO-US05028.
PR 12-MAR-1999; 99US-0123957.
PR 14-MAY-1999; 99US-0134287.
PR 02-JUN-1999; 99WO-US12252.
PR 23-JUN-1999; 99US-0141037.
PR 20-JUL-1999; 99US-0144758.
PR 26-JUL-1999; 99US-0145698.
PR 01-SEP-1999; 99WO-US20111.
PR 08-SEP-1999; 99WO-US20594.
PR 15-SEP-1999; 99WO-US21090.
PR 15-SEP-1999; 99WO-US21547.
PR 05-OCT-1999; 99WO-US23089.
PR 30-NOV-1999; 99WO-US28313.
PR 30-NOV-1999; 99WO-US28409.
PR 02-DEC-1999; 99WO-US28564.
PR 02-DEC-1999; 99WO-US28565.
XX (GETH ) GENENTECH INC.
XX PI Ashkenazi AJ, Baker KP, Ferrara N, Gerber H, Goddard A;

```



Qy	250	actgctgcgtaaagacattccagctttaccaccaatgggttggtgggattctcccgaatggtaga	305
Db	241	acigcgtctaaagacattccagctttaccaccaatgggttggtgggattctcccgaatggtaga	300
Qy	310	gtatgctccctgagatggaatcagcttgagctctctgcgaattggttcacaaactattcatgct	369
Db	301	gtgtgctccctgagatggaatcagcttgagctctctgcgaattggttcacaaactattcatgct	360
Qy	370	tctgtgattctccaaactacttaccttgcctcagatataccctttatctcttaatacagt	429
Db	361	tctgtgattctccaaactacttaccttgcctcagatataccctttatctcttaatacagt	420
Qy	430	ttattttcttcgaataaaaaataactatgagcaacaaaaaataaaaaa	480
Db	421	ttatttttctcgaataaaaaataactatgagcaacataaaaaaataaaaaa	471

RESULT 15

AAV31992

ID AAV31992 standard; DNA; 482 BP.

XXXXXX

AC AAV31992;

XX

DT 25-SEP-1998 (first entry)  
 XX  
 XX BS106 polynucleotide consensus sequence.  
 DE  
 XX ss: human: BS106: breast tissue gene: breast cancer: detection marker.  
 KW

Query Match	96.9%	Score 465.2;	DB 19;	Length 482;
Best Local Similarity	99.6%;	Pred. No. 8.8e-106;		
Matches 464;	Conservative	2;	Mismatches 0;	Indels 0;

QY	1	gccttgaagcattttt9tctgtgcccctgattcttcaggtccaccaccatgaagtctctta	60
DB	16	ggcttgaagcattttt9tctgtgcccctgattcttcaggtccaccaccatgaagtctctta	75
QY	61	gcaccccccgtacctctcttggaagtttccattctctgtgtctctgccagaaatccgaacaa	120

Db	76	gcagctcctggtaactcttggagattccatctttctggtctctgccagaaatccgacaaca	135
Qy	121	gctgctccagctgacacgtaatccagctactggtccttgctgatagtgaagccctcatgct	180
Db	136	gctgctccagctgacacgtaatccagctactggtccttgctgatagtgaagccctcatgct	195
Qy	181	gaaaccaactgctgtaaacccactcgagacacgtgctgacctaccactgcaaccaccgct	240
Db	196	gaaaccaactgctgtaaacccactcgagacacgtgctgacctaccactgcaaccaccgct	255
Qy	241	gcttctaccactgctgtaaaagacattccagttttaccacaaatgggttggggatctcccg	300
Db	256	gcttctaccactgctgtaaaagacattccagttttaccacaaatgggttggggatctcccg	315
Qy	301	aatggttagagtgtgctcctgagatggaatcaggttagtcttctgcgaattgggcacaact	360
Db	316	aatggttagagtgtgctcctgagatggaatcaggttagtcttctgcgaattgggcacaact	375
Qy	361	attcatgcttctgtagtttcatccaaactacttacccttgctctacgataccccctttatct	420
Db	376	attcatgcttctgtagtttcatccaaactacttacccttgctctacgataccccctttatct	435
Qy	421	ctaacagttttattttcttccaataaaaaataactatgagcaaca	466
Db	436	ctaacagttttattttcttccaataaaaaataactatgagcaaca	481

Search completed: April 29, 2002, 19:50:01  
Job time: 22399 sec





strand cDNA was primed with a Not I - oligo(dT) primer [5', AACTGGAAGATTCGGCGCAATATTTTITTTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization. Library constructed by Bento Soares and M. Fatima Bonaldo.

89 a 95 c 108 g 108 t

Query Match 100.0%; Score 382; DB 9; Length 400;  
Best Local Similarity 100.0%; Pred. No. 1.4e-67;  
Matches 382; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 agcgggagcagaagctaaagcccaagcccaagagtgccagtgccagcactgggtgccca 60  
|||||  
Db 383 AGCGGGAGCAGAGCTAAGCCAAAGCCCAAGAGAGTGCCAGTGCCAGCCTGGTGCCA 324  
|||||

QY 61 gtaccagtcaccataaacagtgccagtgccagtgccagcaccagtggtgcttcagtgctg 120  
|||||  
Db 323 GTACCACTACCAATAACAGTGCCAGTGCCAGTGCCAGCAGCAGTGCTGAGTCTG 264  
|||||

QY 121 gtaccagtcaccagcagcactctcacatttggtgctcttcgctggccttggtgagctgg 180  
|||||  
Db 263 GTCCAGCCCTGACCGCCACTCTCACATTGGGCTCTTCGCTGGCCCTTGGTGGAGCTGG 204  
|||||

QY 181 ccagcaccagtgccagctctggtgctggtgttcttcctcacaaagtgcagatttttagatatt 240  
|||||  
Db 203 CCAGCACCAGTGCCAGCTGTGGTGGCTGTCTCTCTCAAGTGAGATTTTAGATAAT 144  
|||||

QY 241 gtaactctccagctcttctcttccttcagcagtggtgcatactcctcagaaactcctcaacac 300  
|||||  
Db 143 GTTAATCCTCCAGCTCTTCTCTCAAGCCAGGTCATCTCTCAGAAACCTACTCAACAC 84  
|||||

QY 301 agcactctaggcagcactatcaatcaattgaagtgcacactctgcattaaactctattg 360  
|||||  
Db 83 AGCACTCTAGCAGCCACTATCAATCAATTGAAGTTGACACTCTGCATTAAATCTATTG 24  
|||||

QY 361 ccattaaaaaahhh 382  
|||||

Db 23 CCATTAAAAAHH 2

RESULT 2  
AA535894/c  
LOCUS  
DEFINITION  
n94a08.s1 NCI\_CGAP\_Co3 Homo sapiens cDNA clone IMAGE:927542 3', similar to contains element TAR1 repetitive element ;, mRNA sequence.  
ACCESSION  
AA535894  
VERSION  
AA535894.1 GI:2280147  
KEYWORDS  
EST.  
SOURCE  
human.  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1 (bases 1 to 535)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Email: [cgaps-r@mail.nih.gov](mailto:cgaps-r@mail.nih.gov)  
Tissue Procurement: Elias Campo, M.D., Michael R. Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: M. Bento Soares, Ph.D.  
CDNA Library Arraying: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LIN at: [www-bio.llnl.gov/bbrp/image/image.html](http://www-bio.llnl.gov/bbrp/image/image.html)  
Insert Length: 584 Std Error: 0.00

Seq primer: -t0m13 fwd. Et from Amersham  
High quality sequence stop: 269.

FEATURES  
Source

Location/Qualifiers  
1..535  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:927542"  
/clone\_lib="NCI\_CGAP\_Co3"  
/sex="pooled"  
/tissue\_type="colon"  
/lav\_host="DH108"

Note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA was prepared from 12 pooled bulk tumor samples and primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization."

BASE COUNT 112 a 129 c 150 g 144 t  
ORIGIN

Query Match 100.0%; Score 382; DB 9; Length 535;  
Best Local Similarity 100.0%; Pred. No. 1.3e-67;  
Matches 382; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 agcgggagcagaagctaaagcccaagcccaagagtgccagtgccagcactgggtgccca 60  
|||||  
Db 386 AGCGGGAGCAGAGCTAAGCCAAAGCCCAAGAGAGTGCCAGTGCCAGCCTGGTGCCA 327  
|||||

QY 61 gtaccagtcaccataaacagtgccagtgccagtgccagcaccagtggtgcttcagtgctg 120  
|||||  
Db 326 GTACCACTACCAATAACAGTGCCAGTGCCAGTGCCAGCAGCAGTGCTTCAGTGCTG 267  
|||||

QY 121 gtaccagtcaccagcagcactctcacatttggtgctcttcgctggtggtggtggtggtg 180  
|||||  
Db 266 GTCCAGCCTGACCCGCACTCTCACATTGGGCTCTTCGCTGGCCCTGGTGGAGCTGGTG 207  
|||||

QY 181 ccagcaccagtggtgctctggtgctggtgttcttcctcacaaagtgcagatttttagatatt 240  
|||||  
Db 206 CCAGCACCAGTGCCAGCTGTGGTGGCTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 147  
|||||

QY 241 gtaactctgccaatttcttctcttcagcaggtgcacactcctcagaaactcctcaacac 300  
|||||  
Db 146 GTTAATCTCTCCAGCT 87  
|||||

QY 301 agcactctaggcagcactatcaatcaattgaagtgcacactctgcattaaactctattg 360  
|||||  
Db 86 AGCACTCTAGCAGCAGCAGTATCAATCAATTGAAGTTGACACTCTGCATTAAATCTATTG 27  
|||||

QY 361 ccattaaaaaahhh 382  
|||||  
Db 26 CCATTAAAAAHH 5

RESULT 3  
BG939439/c  
LOCUS  
DEFINITION  
cr54g11.x1 J1a bone marrow stroma Homo sapiens cDNA clone  
HBN5C\_cr54g11.3', mRNA sequence.

ACCESSION  
BG939439  
VERSION  
BG939439.1 GI:14338811  
KEYWORDS  
EST.  
SOURCE  
human.

ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1 (bases 1 to 699)  
Jia, L.B., You, M.F., Touchman, J.W., Bouffard, G.G., Beckstrom-Sternberg, S.M., Green, E.D., Powell, J.L., Yang, L.M., Rorby, P.G., Hotchkiss, R.N. and Francomano, C.A.  
SGAP: The Skeletal Genome Anatomy Project







D	b	265	GTGGCAGCCCTGACCGCACTCTCACATTTGGGCTTCCTCGCTGCCCTTGGTGAGCTGGTG	206
Q	y	181	ccagcacagtgcagctcgtgctggctggtttcctcaacagtgcagattttagatatt	240
D	b	205	CCACACAGTGGCAGCTTTGGTGCCTCTGGTTTCTCTACAAGTGAATTTTAGATAIT	146
Q	y	241	gttaactcgtccagttctttctctcaagccagggtgcatactcagaaccactactcaaac	300
D	b	145	GTTAATCTGCCAGTCTTTCTCTTAAGCCAGGGTGCATCTCTCAGAAACCCTACTCAAC	86
Q	y	301	agcactcagcgacctcatcaatcaattgaagtgcacctctgcattaaatctattg	360
D	b	85	AGCACTCTAGCAGCCACTATCAATCAATTGAAGTTGACACCTGCATTAATTAATCTIAT	26
Q	y	361	ccattaaaaaaaahhhhhhhh 382	
D	b	25	CCATTTCAAAAAAAAHHHHHHH 4	
RESULT 7				
A1921707/c				
LOCUS				
DEFINITION				
w029db10.x1 NCI_CGAP_Gas4 Homo sapiens cDNA clone IMAGE:2456731 3'				
similar to TR:O76058 O76058 DJ1405.2 ;contains element LIR5				
repetitive element ;; mRNA sequence.				
A1921707				
A1921707.1 GI:5657671				
EST.				
SOURCE				
human.				
ORGANISM				
Homo sapiens				
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
1 (bases 1 to 680)				
NCI-CGAP http://www.ncbi.nlm.nih.gov/hicgap.				
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),				
Tumor Gene Index				
Unpublished (1997)				
JOURNAL		Contact:	Robert Strausberg, Ph.D.	
COMMENT		Email:	cgapbs-remail.nih.gov	
		Tissue Procurement:	Christopher Moskaluk, M.D., Ph.D., Michael R.	
		Emmert-Buck, M.D., Ph.D.		
		cDNA Library Preparation:	Life Technologies, Inc.	
		cDNA Library Arrayed by:	Greg Lennon, Ph.D.	
		Cloning Sequencing by:	Washington University Genome Sequencing Center	
		Clone distribution:	NCI-CGAP clone distribution information can be	
		found through the I.M.A.G.E. Consortium/LINL at:		
		www.bio.lnl.gov/bbrp/image/image.html		
		Insert length:	1142 Std Error: 0.00	
		Seq primer:	-400P from Gibco	
		High quality sequence stop:	423.	
Location/Qualifiers				
1..680				
/organism="Homo sapiens"				
/db_xref="taxon:9606"				
/clone=IMAGE:2456731				
/tissue_type="poorly differentiated adenocarcinoma with				
signed ring cell features"				
/lab_host="DH10B"				
/note="Organ: stomach; Vector: pCMV-SPORT6; Site.1: Salt;				
Site.2: NotI; Cloned unidirectionally. Primer: Oligo dT.				
Average insert size 1.65 kb. Life Technologies catalog #:				
11349-011				
BASE COUNT		143 a	184 c	176 g 176 t 1 others
ORIGIN				
Query Match 98.6%; Score 376.8; DB 9; Length 680;				
Best Local Similarity 99.5%; Pred. No. 1.4e-66;				
Matches 378; Conservative 0; Mismatches 2; Indels 0; Gaps 0;				
Qy		1	aggcgggagcagaagctaaagccaaagccccaaagagtgccagtcgccagcactggtgcaca	60

2283

Query Match 98.2%; Score 375; DB 9; Length 567;  
Best Local Similarity 100.0%; Pred. No. 3.4e-66;  
Matches 375; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 agcggagcagaagctaaagcccaagcccaagagagtgccagtgccagcactgtgcca 60
    |||
Db 375 AGCGGAGCAGAGCTAAAGCCAAAGCCAAAGAGAGTGCGAGTGCCAGCAGCTGTGCCA 316
    |||
QY 61 gtaccagtaccataaacagtgccagtgccagtgccagcaccagtggtgcttcagtgctg 120
    |||
Db 315 GTACCAGTACCATAACAGTGCCAGTGCCAGTGCCAGTGCCAGCAGCAG1GCTGCTTCAGTGTG 256
    |||
QY 121 gtccagcagtcgacgcgaactctcaaatgtggctgttggctgttgcgtgcttggtagagctgctg 180
    |||
Db 255 GTCCAGCCTGACGCGGCACTCTCACATTTGGGCTCTTCCTGCTGGCTTGGTGAGCTGGTG 196
    |||
QY 181 ccagcaccagtgccagctctgtgctgtggtttctctctcaccagtgagattttagatat 240
    |||
Db 195 CCAGCACCAGTGCCAGCTCTGTGTGGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 136
    |||
QY 241 gtaactcctcagctgttctctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 300
    |||
Db 135 GTTAATCCTCCAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 76
    |||
QY 301 agcactctagcagcagcactcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 360
    |||
Db 75 AGCACTTAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 16
    |||
QY 361 ccattataaaaaaa 375
    |||
Db 15 CCATTATAAAAAAAA 1
    |||
```

RESULT 9  
AA971201/c  
LOCUS  
DEFINITION op06all.s1 NCI\_CGAP\_Kid6 Homo sapiens CDNA clone IMAGE:1574876 3', mRNA sequence.  
ACCESSION AA971201  
VERSION AA971201.1 GI:3146491  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
AUTHORS 1 (bases 1 to 383)  
TITLE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
JOURNAL National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
COMMENT Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Email: cgaps-r@mail.nih.gov  
Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R. Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: Stratagene, Inc.  
DNA Sequencing by: Greg Lennon, Ph.D.  
Cloning Distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www.bio.llnl.gov/borpp/image/image.html  
Seq primer: -40ml3 fwd. Et from Amersham  
High quality sequence stop: 369.  
Location/Qualifiers 1..383  
FEATURES  
source  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_image="1574876"  
/clone\_id="NCI-CGAP\_Kid6"  
/sex="mixed"  
/tissue\_type="kidney tumor"  
/lab\_host="SOLR (kanamycin resistant)"  
/note="Organ: kidney; Vector: Bluescript SK-; Site\_1:

ECOR1: Site 2: XhoI; Cloned unidirectionally. Primer:  
Oligo dT. Pooled kidney tumors. 5' adaptor sequence: 5',  
GAATTCGGCAGCAG 3' 3' adaptor sequence: 5',  
CTCGAGTCTTTTCTTTTCTTTT 3' Average insert size: 1.0 kb.\*

BASE COUNT 90 a 90 c 102 g 101 t  
ORIGIN

Query Match 97.8%; Score 373.6; DB 9; Length 383;  
Best Local Similarity 98.9%; Pred. No. 7.2e-66;  
Matches 376; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```
QY 3 gcggagcagagctaaagcccaagcccaagagtgccagtgccagcactgtgctgcca 62
    |||
Db 363 GCGGAGCAGAGCTAAAGCCAAAGCCCAAGAGAGTGCCAGTGCCAGCAGCTGTGCCAGT 324
    |||
QY 63 accagtaccataaacagtgccagtgccagtgccagcaccagtggtgcttcagtgctgctg 122
    |||
Db 323 ACCAGTACCAATAACAGTGCCAGTGCCAGTGCCAGTGCCAGCAGCAGTTCAGTGTGCT 264
    |||
QY 123 gccagcctgagcgcgaactctcactttgggctcttgcctgctgctgctgctgctgctgctg 182
    |||
Db 263 GCCAGCCTGACCGCCACTCTCACATTTGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 204
    |||
QY 183 agcaccagtgccagctgtgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 242
    |||
Db 203 AGCACAGTGCCAGCTCTGGTGGCTGTGCTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 144
    |||
QY 243 laactctgccagctcttctctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 302
    |||
Db 143 TAACTCTGCCAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 84
    |||
QY 303 cactctagcagcagcactatcaatcaatgaagtgaagtgaagtgaagtgaagtgaagtgaagtga 362
    |||
Db 83 CACTCTAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 24
    |||
```

QY 363 attataaaaaaa 382  
Db 23 ATTCTGAAAAA 4  
REFLUT 10  
AA703778/c  
LOCUS  
DEFINITION AA703778 394 bp mRNA linear EST 24-DEC-1997  
IMAGE:1140976 5', mRNA sequence.  
ACCESSION AA703778  
VERSION AA703778.1 GI:2713656  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
AUTHORS 1 (bases 1 to 394)  
TITLE Hillier, L., Allen, M., Bowles, L., Dubuque, T., Gaisel, G., Jost, S.,  
J., Moore, B., Schellenberg, K., Lacy, M., Le, N., Lennon, G., Marra, M., Martin,  
White, Y., Wyllie, T., Waterston, R., Waterston, R., Tan, F., Theising, B.,  
WashU-MCI human EST Project  
Unpublished (1997)  
Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
This clone is available royalty-free through LLNL: contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Seq primer: -28ml3 revl Et from Amersham  
High quality sequence stop: 305.  
Location/Qualifiers 1..394  
FEATURES  
source  
/organism="Homo sapiens"

```

/clone="IMAGE:1140976"
/clone_lib="Stratagene hNT neuron (#937233)"
/dev_stage="hNT neurons"
/lab_host="SOLR (kanamycin resistant)"
/notes="Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI; Cloned unidirectionally, Primer: Oligo dT. Insert differentiated, post mitotic hNT neurons. Average insert size: 1.5 kb; Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGGACGAG 3' -3' adaptor sequence: 5' CTCAGTGTGTGTGTGTGTGT 3'."
CTCAGTGTGTGTGTGTGT 3'."
BASE COUNT      91 a  94 c  109 g  100 t
ORIGIN

Query Match      97.5%; Score 372.6; DB 9; Length 394;
Best Local Similarity 98.9%; Pred. No. 1.le-65;
Matches 375; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 agcgagagcagaagctaaagccaaagcccaagagagtgccagtcgagcactggtgcc 60
    |||||
Db 379 AGCGGGAGCAGACAGCTTAAGCCCAAGCCCAAGAGAGTGCCAGTGCACACCTGGTCCCA 320

Qy 61 gtaccagtcaccataaagcagtgccagtgccagtcgagcagtcgagtcgagtcg 120
    |||||
Db 319 GTACCAAGTACCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAG 260

Qy 121 gtgcagcctgacgcagcactcactcttggctcttcgctgctgctggtggagctgg 180
    |||||
Db 259 GTGCCAGCTGACGCCACTCTCACATTTGGCTCTTCGCTGGCTGGCTGGCTGGCT 200

Qy 181 ccagcagcagtcgagcctggtgctggtgctggtgctggtgctggtgctggtgctg 240
    |||||
Db 199 CCAGCAGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAG 140

Qy 241 gttactctgcagctcttcttcaagcagagtgatctcagaaacctactcaacac 300
    |||||
Db 139 GTTAATCTGCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 80

Qy 301 agcactctgagcagcactcaatcaattgaattgacacactctgcattaaactattg 360
    |||||
Db 79 AGCACTCTAGCAGCAGTATCAATCAATGAAGTTGACACTCTGCATTAATTAATTA 20

Qy 361 ccattaaaaaataaaaaa 379
    |||||
Db 19 CCATTTCTGAAAAAATAAAA 1

RESULT 11
LOCUS      BF196380
DEFINITION 7n69d03.x1 NCI_CGAP_Ov18 Homo sapiens cDNA clone IMAGE:3569717 3'
            similar to TR:076058 076058 DJ1409.2 :contains element MER1
            repetitive element ;; mRNA sequence.
ACCESSION  BF196380
VERSION     BF196380.1 GI:11084264
KEYWORDS   EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
            1 (bases 1 to 684)
            NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
            National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
            Unpublished (1997)
            Contact: Robert Strausberg, Ph.D.
            Email: cgapsb-r@mail.nih.gov
            Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
            R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: M. Bento
            Soares, Ph.D. cDNA Library Arrayed by: Christa Prange, Ite
            I.M.A.G.E. Consortium DNA Sequencing by: Washington University
            Genome Sequencing Center

```

Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL, send email to: i:fo:image.llnl.gov

High quality sequence stop: 389.

Location/Qualifiers

source

1..684

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:3569717"

/clone\_lib="NCI\_CGAP\_Ov18"

/tissue\_type="fibrothoma"

/lab\_host="DH10B (phage-resistant)"

/note="Organ: ovary; Vector: pi713p-Pac (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' IGTTACCAATCTGAAGTGGAGCGCGCGACATTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pi713 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 151 a 179 c 176 g 177 t

ORIGIN

Query Match 97.5%; Score 372.6; DB 10; Length 684;

Best Local Similarity 98.9%; Pred. No. 9.8e-66;

Matches 375; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 agcgagagcagaagctaaagccaaagcccaagagagtgccagtcgagcactggtgcc 60

Db 379 AGCGGGAGCAGACAGCTTAAGCCCAAGCCCAAGAGAGTGCCAGTGCACACCTGGTCCCA 320

Qy 61 gtaccagtcaccataaagcagtgccagtgccagtcgagcagtcgagtcgagtcg 120

Db 319 GTACCAAGTACCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAG 260

Qy 121 gtgcagcctgacgcagcactcactcttggctcttcgctgctgctggtggagctgg 180

Db 259 GTGCCAGCTGACGCCACTCTCACATTTGGCTCTTCGCTGGCTGGCTGGCTGGCT 200

Qy 181 ccagcagcagtcgagcctggtgctggtgctggtgctggtgctggtgctggtgctg 240

Db 199 CCAGCAGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAG 140

Qy 241 gttactctgcagctcttcttcaagcagagtgatctcagaaacctactcaacac 300

Db 139 GTTAATCTGCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 80

Qy 301 agcactctgagcagcactcaatcaattgaattgacacactctgcattaaactattg 360

Db 79 AGCACTCTAGCAGCAGTATCAATCAATGAAGTTGACACTCTGCATTAATTAATTA 20

Qy 361 ccattaaaaaataaaaaa 379

Db 19 CCATTTCTGAAAAAATAAAA 1

RESULT 12

LOCUS AI091673

DEFINITION Oo23e09.x1 Soares\_NSF\_F8\_9W\_OT\_PA\_P\_S1 Homo sapiens cDNA clone

IMAGE:1567048 3', mRNA sequence.

ACCESSION AI091673

VERSION AI091673

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 425)

NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Email: cgapsb-r@mail.nih.gov

Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael

R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: M. Bento

Soares, Ph.D. cDNA Library Arrayed by: Christa Prange, Ite

I.M.A.G.E. Consortium DNA Sequencing by: Washington University

Genome Sequencing Center









GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 29, 2002, 17:17:18 ; Search time 6783.7 seconds

(without alignments)

1144.029 Million cell updates/sec

Title:

US-09-248-178-65

Perfect score:

575

Sequence:

1 accatataataaaaggatat.....aaaggataaaaaaaaaaaaaa 575

Scoring table:

IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched:

13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters:

27472414

Minimum DB seq length:

0

Maximum DB seq length:

2000000000

Post-processing:

Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*

1: em\_estba:.\*  
2: em\_esthum:.\*  
3: em\_estin:.\*  
4: em\_estnu:.\*  
5: em\_estov:.\*  
6: em\_estpl:.\*  
7: em\_estro:.\*  
8: em\_hic:.\*  
9: gb\_estl:.\*  
10: gb\_estc2:.\*  
11: gb\_hic:.\*  
12: gb\_gss:.\*  
13: em\_gss\_hum:.\*  
14: em\_gss\_inv:.\*  
15: em\_gss\_pln:.\*  
16: em\_gss\_vrt:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	522.2	90.8	947	9	AL133916
2	519	90.3	757	9	AW955684
3	504	87.7	608	9	AW950828
4	498.6	86.7	549	9	AI346341
5	497	86.4	541	9	AI867454
6	488	84.9	532	10	BM263376
7	423	73.6	467	10	BF432231
8	410	71.3	454	9	AI655270
9	410	71.3	455	9	AI421279
10	405	70.4	449	10	BF439949
11	402	69.9	454	9	AW014882
12	397.6	69.1	486	10	NS4784
13	384.8	66.9	432	9	AA775552
14	366	63.7	410	9	AA626243
15	336	58.4	380	9	AI341407
16	334.2	58.1	423	10	N62351
17	328.4	57.1	409	10	NS9253

BE175639	RC5-HT058	BE175639	422	9	BE175639
BF114841	7i87c07.x	BF114841	404	10	BF114841
AA456968	aa90905.s	AA456968	294	9	AA456968
AA364013	EST74581	AA364013	324	9	AA364013
AI743400	wg92906.x	AI743400	430	9	AI743400
AA457077	aa90905.r	AA457077	277	9	AA457077
AI356918	gy19c12.x	AI356918	276	9	AI356918
N76721	y282b10.r1	N76721	463	10	N76721
AI743400	wg92906.x	AI743400	430	9	AI743400
BE102996	UI-R-BT1-	BE102996	415	9	BE102996
AI577685	UI-R-GO-u	AI577685	352	9	AI577685
AW534413	UI-R-C4-a	AW534413	360	9	AW534413
AA585385	am81903.s	AA585385	273	9	AA585385
BC231440	na142e02.	BC231440	57	10	BC231440
AA998336	UI-R-CO-1	AA998336	200	9	AA998336
AL069616	Drosophil	AL069616	1101	12	CNS00EYP
BM353500	lg56a02.x	BM353500	487	10	BM353500
C90235	C90235 Dict	C90235	517	10	C90235
BI276354	UI-R-CMO-	BI276354	386	10	BI276354
AL067742	Drosophil	AL067742	536	12	CNS00L44
AA714041	nx81908.s	AA714041	334	9	AA714041
C92323	C92323 Dict	C92323	455	10	C92323
AL666548	AL666548	AL666548	758	9	AL666548
C84124	C84124 Dict	C84124	429	10	C84124
BF457592	UI-M-B21-	BF457592	394	10	BF457592
AI104482	Drosophil	AI104482	1067	12	CNS014QG
C92261	C92261 Dict	C92261	564	10	C92261
AL059601	Drosophil	AL059601	868	12	CNS010YV

#### ALIGNMENTS

RESULT 1

AL133916

LOCUS

DEFINITION

AL133916

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

AL133916 947 bp mRNA linear EST 25-FEB-2000  
DKFZP761N0114.1 761 (synonym: hamy2) Homo sapiens cDNA clone  
DKFZP761N0114.5', mRNA sequence.

AL133916.1 GI:6602103

EST.

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 947)

Ansorge, D., Winkler, U., Mewes, H.W., Weil, B. and Wiemann, S.

EST (Ansorge, D., Winkler, U., Mewes, H.W., Weil, B. and Wiemann, S.)

Unpublished (1999)

Contact: Ansorge W

MIPS

Am Klopierseil, 18a D-82152 Martinsried, Germany

This is the 5' sequence of the clone insert

Clone from S. Wiemann, Molecular Genome Analysis, German Cancer

Research Center (DKFZ); Email: s.wiemann@dkfz-heidelberg.de;

Sequenced by EMBL (European Molecular Biology Laboratory);

Heidelberg/Germany) within the cDNA sequencing consortium of the

German Genome Project.

No sl sequence available

This clone (DKFZP761N0114) is available at the RZPD in Berlin.

Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059

Berlin-Charlottenburg, Germany; Email: clone@rzpd.de.

Location/Qualifiers

1..947

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="DKFZP761N0114"

/clone\_lib="761 (synonym: hamy2)"

/cissue\_type="amygdala"

/dev\_stage="adult"

/lab\_host="DH10B"

/note="Vector: pSPORT1; Site\_1: NotI; Site\_2: SalI"

313 170 C 182 G 282 T

BASE COUNT

SEQ ID NO: 65

221

221

221

221

Seq primer: Reverse.  
Location/Qualifiers  
1..757  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="MAGE resequences, MAGD"  
/note="Vector: pBluescriptSkm"  
BASE COUNT 240 a 161 c 146 g 210 t  
ORIGIN  
Query Match 50.3%; Score 519; DB 9; Length 757;  
Best Local Similarity 98.4%; Pred. No. 3.4e-93;  
Matches 566; Conservative 0; Mismatches 5; Indels 4; Gaps 4;  
QY 1 actlgatataaaaggatccataatgaatttttatactgcattcttaccattacattgcca 60  
DB 59 actlgatataaaaggatccataatgaatttttatactgcattcttaccattacattgcca 118  
QY 61 ctataatcgtttattgcttgatgaagacctttccacagaatccctatggtgagcatttca 120  
DB 119 ctataatcgtttattgcttgatgaagacctttccacagaatccctatggtgagcatttca 178  
QY 121 ctggtctacttcacacccatgctttaaagagggcagtttctcctcaaaagcagaacatgccc 180  
DB 179 ctggtctacttcacacccatgctttaaagagggcagtttctcctcaaaagcagaacatgccc 238  
QY 181 gccagttctcaagtgtttctcctcaactccatttgaatglaagggcagctggcccccaatg 240  
DB 239 gccagttctcaagtgtttctcctcaactccatttgaatglaagggcagctggcccccaatg 298  
QY 241 tgggaggtccgaacattttctgaattccattttcttctgctgcgctgaatgacagttt 300  
DB 299 tgggaggtccgaacattttctgaattccattttcttctgctgcgctgaatgacagttt 358  
QY 301 ctgctattactgagttcccgatctttcccaagggtgtgtgatttcaaaagggccagcta 360  
DB 359 ctgctattactgagttcccgatctttcccaagggtgtgtgatttcaaaagggccagcta 417  
QY 361 atagccagaaatcatgacctgaagagagatgaaatttcaagcgtgagccagcagga 420  
DB 418 atagccagaaatcatgacctgaagagagatgaaatttcaagcgtgagccagcagga 475  
QY 421 gctccagtagtgcgaaggttcttgagaatcagccatttgggtcaaaaaagatttttaag 480  
DB 476 gctccagtagtgcgaaggttcttgagaatcagccatttgggtcaaaaaagatttttaag 534  
QY 481 cttttatgttatccatgagccatagaaaggtatggtatgttttaagaactattttaa 540  
DB 535 cttttatgttatccatgagccatagaaaggtatggtatgttttaagaactattttaa 594  
QY 541 gctttccagaccccaaaaggaagaaaaa 575  
DB 595 gctttccagaccccaaaaggaagaaaaa 629  
RESULT 3  
AW955684/c  
LOCUS  
DEFINITION  
EST362896 MAGE resequences, MAGD Homo sapiens cDNA, mRNA sequence.  
ACCESSION  
AW950828  
VERSION  
AW950828.1 GI:8140485  
KEYWORDS  
EST.  
SOURCE  
human.  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 608)  
REFERENCE  
Hegde,P., Qi,X., Abernathy,K., Dharap,S., Gaspar,R., Gay,C., Holt  
J.E., Saeed,A.I., Sharov,V., Lee,N.H., Yeatman,T.J. and  
Quackenbush,J  
Assessment of gene expression patterns in a model of colon tumor  
metastasis using a 19,200 element cDNA microarray

Query Match 90.8%; Score 522.2; DB 9; Length 947;  
Best Local Similarity 98.8%; Pred. No. 7.4e-94;  
Matches 568; Conservative 0; Mismatches 3; Indels 4; Gaps 4;  
QY 1 actlgatataaaaggatccataatgaatttttatactgcattcttaccattacattgcca 60  
DB 241 actlgatataaaaggatccataatgaatttttatactgcattcttaccattacattgcca 300  
QY 61 ctataatcgtttattgcttgatgaagacctttccacagaatccctatggtgagcatttca 120  
DB 301 ctataatcgtttattgcttgatgaagacctttccacagaatccctatggtgagcatttca 360  
QY 121 ctggtctacttcacacccatgctttaaagagggcagtttctcctcaaaagcagaacatgccc 180  
DB 361 ctggtctacttcacacccatgctttaaagagggcagtttctcctcaaaagcagaacatgccc 420  
QY 181 gccagttctcaagtgtttctcctcaactccatttgaatglaagggcagctggcccccaatg 240  
DB 421 gccagttctcaagtgtttctcctcaactccatttgaatglaagggcagctggcccccaatg 480  
QY 241 tgggaggtccgaacattttctgaattccattttcttctgctgcgctgaatgacagttt 300  
DB 481 tgggaggtccgaacattttctgaattccattttcttctgctgcgctgaatgacagttt 540  
QY 301 ctgctattactgagttcccgatctttcccaagggtgtgtgatttcaaaagggccagcta 360  
DB 541 ctgctattactgagttcccgatctttcccaagggtgtgtgatttcaaaagggccagcta 595  
QY 361 atagccagaaatcatgacctgaagagagatgaaatttcaagcgtgagccagcagga 420  
DB 600 atagccagaaatcatgacctgaagagagatgaaatttcaagcgtgagccagcagga 657  
QY 421 gctccagtagtgcgaaggttcttgagaatcagccatttgggtcaaaaaagatttttaag 480  
DB 658 gctccagtagtgcgaaggttcttgagaatcagccatttgggtcaaaaaagatttttaag 716  
QY 481 cttttatgttatccatgagccatagaaaggtatggtatgttttaagaactattttaa 540  
DB 717 cttttatgttatccatgagccatagaaaggtatggtatgttttaagaactattttaa 776  
QY 541 gctttccagaccccaaaaggaagaaaaa 575  
DB 777 gctttccagaccccaaaaggaagaaaaa 811  
RESULT 2  
AW955684  
LOCUS  
DEFINITION  
EST367754 MAGE resequences, MAGD Homo sapiens cDNA, mRNA sequence.  
ACCESSION  
AW955684  
VERSION  
AW955684.1 GI:8145367  
KEYWORDS  
EST.  
SOURCE  
human.  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 757)  
REFERENCE  
Hegde,P., Qi,X., Abernathy,K., Dharap,S., Gaspar,R., Gay,C., Holt  
J.E., Saeed,A.I., Sharov,V., Lee,N.H., Yeatman,T.J. and  
Quackenbush,J  
Assessment of gene expression patterns in a model of colon tumor  
metastasis using a 19,200 element cDNA microarray  
Unpublished (2000)  
Contact: John Quackenbush  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 3528  
Fax: 301 838 0208  
Email: johnq@tigr.org  
Plate: 82

JOURNAL Unpublished (2000)  
COMMENT Contact: John Quackenbush  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 3528  
Fax: 301 838 0208  
Email: johnq@tigr.org  
Plate: 19  
Seq primer: Forward.  
Location/Qualifiers  
1. 608  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="MAGE resources, MAGE"  
/note="vector: paluescriptskm"  
BASE COUNT 184 a 115 c 123 g 186 t  
ORIGIN  
Query Match 87.7%; Score 504; DB 9; Length 608;  
Best Local Similarity 99.3%; Pred. No. 3.4e-90;  
Matches 548; Conservative 0; Mismatches 0; Indels 4; Gaps 4;  
QY 1 acctgataaaagagatccataatgaataattttatctatgcacccctttacattagcca 60  
|||||  
DB 548 ACTTGATATAAAGAGATATCCATATGAATATTTATCTGATCCCTTTACATTAGCCA 489  
QY 61 ctataacgttattgcttgatgaagaccctttcacagaatccctatgattgcagacttca 120  
|||||  
DB 488 CTAATACGTTATTGCTTGATGAAGACCTTTCACAGAAATCCTATGATGTCAGCAATTC 429  
QY 121 ctggctacttataccatgccttlaaagagggcagttcttccaaagcagaacaatgcc 180  
|||||  
DB 428 CTGGCTACTTCTATACCATGCTTAAGAGAGGGCAGTTTCTCAAGACAGAAATGCC 369  
QY 181 gccagtctcaagtttctctcctaaactccatttgaatgtaagggcagctggcccccaatg 240  
|||||  
DB 368 GCCAGTTCTCAAGTTTTCCTCTAACTCCATTTGAATGTAAGGGCAGCTGGCCCCCAAG 309  
QY 241 tggggaggtcccaaacatttctgaaatccatttcttctgttcggcgctaaatgcagttt 300  
DB 308 IGGGAGGTCCGAAACATTTCTGAATTCCTATTTCTGTCGGGCTAAATGACAGATT 249  
QY 301 ctgtcattacttagatcccgatctttcccaaggtgttgcattacaagagggccagcta 360  
|||||  
DB 248 CTGTCATTACTTAGATT-CCGATCTTTCCTCAAGGTGTGATTACAAAGAGGCCAGCTA 150  
QY 361 atagccagaatacatgacctgaagagagatgaaatttcaagctgtgagccagcagga 420  
|||||  
DB 169 ATAG-CAGAAATCATGACCTGAAAGAGAGATCAAA-TTCAAGCTGTGACCCAGGCAGCA 132  
QY 421 gctccagtagtgcgaaggtttcttgagaatcagccatttgggtacaaaaagatttttaag 480  
|||||  
DB 131 GCT-CAATATGCGAAGGTTCTTGAATCAGCCATTTGTGTACAAAAAGATTTTAAAG 73  
QY 481 cttttatgttatccatggagccatagaagagctatgattgtttaagaactatttttaa 540  
|||||  
DB 72 CTTTATGTTATACCATGGAGCCATGAAAGGCTATGGATTGTTTAAGAACTATTTTTAA 13  
QY 541 gtgtccagacc 552  
|||||  
DB 12 GTGTTCCAGACC 1

JOURNAL Unpublished (2000)  
COMMENT Contact: John Quackenbush  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 3528  
Fax: 301 838 0208  
Email: johnq@tigr.org  
Plate: 19  
Seq primer: Forward.  
Location/Qualifiers  
1. 608  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="MAGE resources, MAGE"  
/note="vector: paluescriptskm"  
BASE COUNT 184 a 115 c 123 g 186 t  
ORIGIN  
Query Match 87.7%; Score 504; DB 9; Length 608;  
Best Local Similarity 99.3%; Pred. No. 3.4e-90;  
Matches 548; Conservative 0; Mismatches 0; Indels 4; Gaps 4;  
QY 1 acctgataaaagagatccataatgaataattttatctatgcacccctttacattagcca 60  
|||||  
DB 548 ACTTGATATAAAGAGATATCCATATGAATATTTATCTGATCCCTTTACATTAGCCA 489  
QY 61 ctataacgttattgcttgatgaagaccctttcacagaatccctatgattgcagacttca 120  
|||||  
DB 488 CTAATACGTTATTGCTTGATGAAGACCTTTCACAGAAATCCTATGATGTCAGCAATTC 429  
QY 121 ctggctacttataccatgccttlaaagagggcagttcttccaaagcagaacaatgcc 180  
|||||  
DB 428 CTGGCTACTTCTATACCATGCTTAAGAGAGGGCAGTTTCTCAAGACAGAAATGCC 369  
QY 181 gccagtctcaagtttctctcctaaactccatttgaatgtaagggcagctggcccccaatg 240  
|||||  
DB 368 GCCAGTTCTCAAGTTTTCCTCTAACTCCATTTGAATGTAAGGGCAGCTGGCCCCCAAG 309  
QY 241 tggggaggtcccaaacatttctgaaatccatttcttctgttcggcgctaaatgcagttt 300  
DB 308 IGGGAGGTCCGAAACATTTCTGAATTCCTATTTCTGTCGGGCTAAATGACAGATT 249  
QY 301 ctgtcattacttagatcccgatctttcccaaggtgttgcattacaagagggccagcta 360  
|||||  
DB 248 CTGTCATTACTTAGATT-CCGATCTTTCCTCAAGGTGTGATTACAAAGAGGCCAGCTA 150  
QY 361 atagccagaatacatgacctgaagagagatgaaatttcaagctgtgagccagcagga 420  
|||||  
DB 169 ATAG-CAGAAATCATGACCTGAAAGAGAGATCAAA-TTCAAGCTGTGACCCAGGCAGCA 132  
QY 421 gctccagtagtgcgaaggtttcttgagaatcagccatttgggtacaaaaagatttttaag 480  
|||||  
DB 131 GCT-CAATATGCGAAGGTTCTTGAATCAGCCATTTGTGTACAAAAAGATTTTAAAG 73  
QY 481 cttttatgttatccatggagccatagaagagctatgattgtttaagaactatttttaa 540  
|||||  
DB 72 CTTTATGTTATACCATGGAGCCATGAAAGGCTATGGATTGTTTAAGAACTATTTTTAA 13  
QY 541 gtgtccagacc 552  
|||||  
DB 12 GTGTTCCAGACC 1

RESULT 4  
AI346341/c  
LOCUS qp50b09.x1 NCI\_CGAP\_C08 Homo sapiens cDNA clone IMAGE:1926125 3',  
DEFINITION mRNA sequence.  
ACCESSION AI346341  
VERSION AI346341.1 GI:4083547  
KEYWORDS EST.  
SOURCE human.

Homo Sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 549)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap/>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
Eickert-Buck, M.D., Ph.D.  
CDNA Library Preparation: M. Bento Soares, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone Distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/[www.bio.lnu.gov/bbrp/image/image.html](http://www.bio.lnu.gov/bbrp/image/image.html)  
Insert length: 550 Std Error: 0.00  
Seq primer: -400P from Gibco  
High quality sequence stop: 446.  
Location/Qualifiers  
1. 549  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1926125"  
/clone\_lib="NCI\_CGAP\_C08"  
/tissue\_type="adenocarcinoma"  
/lab\_host="DH10B"  
/note="Organ: colon; Vector: pT7130-Pac (Pharmacia) with a  
modified polylinker; 1st strand cDNA was prepared from a  
colon adenocarcinoma, and was then primed with a Not I -  
oligo(dT) primer. Double-stranded cDNA was ligated to Eco  
RI adaptors (Pharmacia), digested with Not I and cloned  
into the Not I and Eco RI sites of the modified pT73  
vector. Library is normalized. Library was constructed by  
Bento Soares and M. Fatima Bonaudo.  
BASE COUNT 160 a 107 c 110 g 172 t  
ORIGIN  
Query Match 56.7%; Score 498.6; DB 9; Length 549;  
Best Local Similarity 98.6%; Pred. No. 4.1e-89;  
Matches 549; Conservative 0; Mismatches 4; Indels 4; Gaps 4;  
QY 14 aggtatccataatgaatattttatctatgcacccctttacattagccctaaatcgttat 73  
|||||  
DB 549 AGGATATCCATAATGAATATTTATCTGCAATCCTTTACATAGCCCAATACGTTAT 490  
QY 74 tgcctgatgaagcctttcacagagctcctatggatgcagcattcttctggctacttca 133  
|||||  
DB 489 TCGTTGATGAAGACCTTTCACAGAAATCCTATGGATTCGAGCATTTCTACTTGGCTACTICA 430  
QY 134 taccatgccttaagagggcagtttcttctcaaaagcagaacaatgcgcagcttctcaag 193  
|||||  
DB 429 TACCCATGCCITAAAGAGGGGCAGTTTCTCAAAAGCAGAAACATGCCGCCAGTCTCAAG 370  
QY 164 ttctctctactactccatttgaatgtaagggcagctgcccccaatgtggggaggtccga 253  
|||||  
DB 369 TTTTCCTCCTAACCTCAATTTGAAATGTAAGGCAGCTGCCCCCAATGTGGGAGGTCCGA 310  
QY 254 acattttctgaattcccatcttctgttcgcggctaaatgacagttcttctcattactta 313  
|||||  
DB 309 ACAHTTCTGAATTCCTCATTTCTGTCGGCGCTAAATGACAGTTTCTGTCATTACTTA 250  
QY 314 gattccccgatcttccccaaaggtgttgattttacaaagagggcagctaatagccagaatc 373  
|||||  
DB 249 GATT-CCGATCTTTCCTCAAGGTGTTGATTTACAAAGAGGCCAGCTTAATAG-CAGAAATC 192  
QY 374 atgacctctgaagagagatgaatttcaagctgtgagccagcagggagctccagtatggc 433  
|||||  
DB 191 ATGACCCCTGAAGAGAGAGATGAAGAA-TTCAAGCTGTGAGCCAGCCAGGAGCT-CAGTATGCC 134



National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Email: rstraus@nhi.nih.gov  
Tissue procurement: David N. Louis, M.D., Myrna K. Rosenfeld M.D.,  
Ph.D.  
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima  
Bonaldo, Ph.D.  
cDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL, send email to:  
info@image.llnl.gov  
Seq primer: -400P from Gibco.  
Location/Qualifiers  
1. 167  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="HRR5 Islet"  
/tissue\_type="Purified pancreatic islet"  
/lab\_host="DH10B"  
/Note="Organ: Pancreas; Vector: pBluescript SK(-); Site\_1:  
NotI; Site\_2: XhoI; cDNA made by oligo-dT priming.  
Size selected on agarose gel. Average insert size -1kL. 5'  
XhoI site was destroyed after directional cloning.  
Amplified once. Contact information: Hiroshi Inoue, MD,  
Metabolism Div. (Alan Permutt Lab), Washington University  
School of Medicine, Box 8127, 660 South Euclid Ave., St.  
Louis, MO 63110, E-mail: hinoue@imgate.wustl.edu, tel:  
314-362-1916, fax: 314-747-2692."  
BASE COUNT 158 a 106 c 108 g 160 t  
ORIGIN  
Query Match 84.9%; Score 488; DB 10; Length 532;  
Best Local Similarity 99.3%; Pred. No. 5.2e-87;  
Matches 532; Conservative 0; Mismatches 0; Indels 4; Gaps 4;  
QY 24 taatgaatattttactgcatctttacattagccaataacgtttattgctgatga 83  
DB 532 TAATGAATATTTTATACGCACTCTTACATTAAGCACTAAATACGTTATGCTTATGA 473  
QY 84 agacctttcacagaatcctatggttcgagcatttcacttggctacttcatacccatgcc 143  
DB 472 AGACCTTTACAGAACTCCTATGATTCAGCATTTCACTTGGCTACTTCATACCCATGCC 413  
QY 144 ttaagaggggcagttctctcaaaagcagaaacatgcgcgcagttctcaatgttccctcc 203  
DB 412 TTAAGAGGGGCAGTTTCTCAAAAGCAGAAACATGCCGCCAGTTCTCAAGTTTTCCTCCI 333  
QY 204 aactccattgaatgaaggcagctggcccccataatgaggaggttcggaacattttctg 263  
DB 352 AACTCCATTGAATGAAGGCAGCTGGCCCCCAATGTGGGAGGTCCGAACATTTCIG 293  
QY 264 aattccattttctgttcggttaaatgacagtttctgtcattactttagattcccgat 323  
DB 252 AATTCCATTTCCTGTTTCGGGGTAAATGACAGTTTCGTCACTTACCTTAGATT-CCGAT 234  
QY 324 cttcccaaaagctgtgatttacaagaagccagctaatagccagaatcatgacctga 383  
DB 233 CTTTCCCAAAAGCTGTGATTACAAAGAGGCCAGCTAATAG-CAGAAATCATGACCTGA 175  
QY 384 aagagagatgaatttcacgtgtgagccagcgagggagctccagtcagtgagcaagttctt 443  
DB 174 AACAGAGATGAAA-ATCAAGCTGTGAGCCAGCGAGGAGCT-CAGTAIGGCAAGGTTCTT 117  
QY 444 gagaatacaccatttgggtacaaaagaatttttcaagcttttcaagcttttcaacatggagcc 503  
DB 116 GAGAAATCAGCCATTGGTACAAAAGATTTTAAAGCTTTTATGTTTATACATGGAGCC 57  
QY 504 atagaaggctatggattgttcaagaactattttttaaaggttccagaccocaaaaag 559  
DB 56 ATAGAAAGGCTATGGATTGTTTAAAGAACTATTTTAAAGTGTTCACAGCCCAAAAAG 1

BASE COUNT 134 a 98 c 98 g 137 t  
ORIGIN  
Query Match 73.6%; Score 423; DB 10; Length 467;  
Best Local Similarity 59.2%; Pred. No. 3.9e-74;  
Matches 467; Conservative 0; Mismatches 0; Indels 4; Gaps 4;  
QY 83 aagagcttttcaagaatcctatggttcgagcatttcacttgcctacttcatacccatgc 142  
DB 467 AAGACCTTTCACAGAACTCCTATGCTATGAGCATTTTCAGTTTCAGTTTTCATACCCATGC 408  
QY 143 ctttaagaagctgtgatttctcaaaagcagaataatgcgcgcgttctccagtttctctcc 202  
DB 407 CTTAAAGAGTGGCAGTTTCTCAAAAGCAGAAACATGCCGCCAGTTTCAGTTTTCCTCC 348  
QY 203 taccctccattgaaatgaaaggcagctggcccccataatgaggaggttcggaacattttct 262  
DB 347 TAACTCCATTGAAATGAAGGCGAGCTGGCCCCCAATGTGGGAGGTTCGTAAGATT-CCGA 259  
QY 263 gaattcccatcttctgttcgcggtctaaatgacagtttctgtcattacttagattcccca 322  
DB 287 GAATTCCTCAATTCCTGTTTCGGGCTAAATGACAGTTTCTGCTATGATTAGATT-CCGA 229  
QY 323 tcttccccaaggtgtgtgatttcaaaagaagccagctaatagccagaataatcagaccctg 382  
DB 228 TCTTTCCTCAAGGTTGTTGATTACAAAGAGGCGAGCTAATAG-CAGAAATCATGACCTCG 170  
QY 383 aagagagatgaatttcaagctgtgagccagcgaggtccagtcagtcagtcagtcagtcagtc 442  
DB 169 AAAGAGAGATGAAA-ATCAAGCTGTGAGCCAGCGAGGAGCT-CAGTAIGGCAAGGTTCT 112  
QY 443 tgaagaatcaacatcttgcggtacaaaagaatttttaaagctttttagtttaccatagagc 502  
DB 111 TCAGAAATCAGCCATTGGTACAAAAGATTTTAAAGCTTTTAAAGTGTTCATGATTACCATGAGC 52  
QY 503 catagaagaagcttatgaggttttaagaactattttttaaagtttccagaccoc 553  
DB 51 CATAGAAAGGCTATGGATTGTTTAAAGAACTATTTTAAAGTGTTCACAGCCCAAAAAG 1

RESULT 7  
BF432231/c  
LOCUS nab87a12.x1 NCI-CGAP\_Brn23 Homo sapiens cDNA clone IMAGE:3274655  
DEFINITION 3', mRNA sequence.  
ACCESSION BF432231  
VERSION BF432231.1 GI:11444345  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 467)  
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.





[illegible]



```

|||||
Db 486 ACGTATTGCTGTGATGAAGACCTTTTACNGAATCCCTATGATTTGCGACATTCACATTCG 427
Qy 127 tacttcata-cccatgecttaagaggggagctttctcaaaagcagaacacatgccg-cca 184
|||||
Db 426 TACTTCAATACCCATGCTTTAAAGAGGGGCGCACTTCTCAAAAGCAGAAACATGCCGCCCA 367
Qy 185 gtctcaagtttctctcaactccattgaatgaatgaaggcagctggcccccgaatgtgg 244
Db 366 GTTCTCAAGTTTCTCTCACTCCATTTGAATGTAAGGCGAGCTGCCCCCAATGTGG 307
Qy 245 gaggtccgaaca-tttctgaattcccatctttctgttcgcggcgttaaatgacagtttctg 303
|||||
Db 306 GAGGTCCGGAACATTTTCTGAATTCCTATTTCTGTGTGCGGCTAAATGACAGATTCTG 247
Qy 304 tcattactagattcccgatctttcccaagggtgtgatttacaagaagggccagcttaata 363
Db 246 TCATTACTTAGATT-CCGATCTTTCCCAAGGTTGTGATTTACAAAGAGGCCAGCTAATA 188
Qy 364 gccagaatcatgacctgaaagagagatgaatttcaagctgtgagccagggcagagct 423
Db 187 G-CAGAAATCATGACCTGAAAGAGAGATGAAA-TTCAAGCTGTGACCCAGGCGAGGCT 130
Qy 424 ccagtatggcaagggtttcttgagatcagccatttgggtacaaaaaagatttttaagctt 483
Db 129 -CAGTATGGCAAGGTTCTTGAGAAATCAGCCATTGTGTACAAAAGAGATTTTTAAAGCT 71
Qy 484 ttatgttataccatggagccatgaaggctatgattgttttaagactatttttaagctg 543
Db 70 TTATGTATACCATGGAGCCATGAAGGCTATGATTTGTTAAGAACTATTTTAAAGT 11
Qy 544 ttccagacct 553
Db 10 TTCCAGAGCC 1

RESULT 13
AA775552/c
LOCUS
DEFINITION
IMAGE: 2f25g02.s1 Soares_fetal_heart_NDHH19W Homo sapiens cDNA clone
IMAGE: 378002 3', mRNA sequence.
AA775552
ACCESSION
AA775552.1 GI:2834886
VERSION
AA775552.1
KEYWORDS
EST
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 432)
Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,
Krizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin
J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F., Theising, B.,
White, Y., Wyllie, T., Waterston, R. and Wilson, R.
WashU-NCI human EST Project
Unpublished (1997)
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.wustl.edu
This clone is available royalty-free through LLNL: contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 431.
Location/Qualifiers
1. 432
/organism="Homo sapiens"
/db_xref="GDB:1286258"
/db_xref="taxon:9606"
/clone="IMAGE:378002"
/clone_lib="Soares_fetal_heart_NDHH19W"
/sex="unknown"
FEATURES
Source

```

```

/dev_stage="1y weeks"
/lab_host="DHI0B (ampicillin resistant)"
/notes="Organ: heart; Vector: pT73D (Pharmacia) with a
modified polylinker; Site_1: Not 1; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5.
TGTTACCAATCTGAAGTGGAGCGCGCATTTTTTTTTTTTTTTT 3'].
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT73 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by
M.Fatima Bonaldo. This library was constructed from the
same fetus as the fetal lung library, Soares fetal lung
NBHH19W."
BASE COUNT 122 a 92 c 87 g 131 t
ORIGIN

Query Match. 66.9%; Score 384.8; DB 9; Length 432;
Best Local Similarity 98.6%; Pred. No. 1.4e-66;
Matches 430; Conservative 0; Mismatches 2; Indels 4; Gaps 4;

Qy 126 ctacttcataccatgaccttaagagggcagcttctcaaaagcagaacacatgcccgag 185
Db 432 CTACTTCATACCCATGCCITAAAGAGGGCGAGTTTCTCAAAAGCAGAAACATGCCGCCAG 373
Qy 186 ttctcaagtttctctcaactccatttgaatgaaggcgagctggcccccgaatgtgggg 245
Db 372 TTCICAAGTTTCTCTCACTCACTCAATTTCAATGTAAGGCGAGCTGGCCCATATGTGGG 313
Qy 246 aggtccgaacatttctgaattcccatcttctgttcgcggcgttaaatgacagcttctgc 305
Db 312 AGGTCCGGAACATTTCTGAATTCCTATTTCTGTGTGCGGCTAAATGACAGATTTCGTGC 253
Qy 306 attacttagattcccgatctttcccaagggtgtgatttacaagaagggcgagctaatagc 365
Db 252 ATTACTTAGATT-CCGATCTTTCCCAAGGTTGTGATTTACAAAGAGCGCGAGCTAATAG- 193
Qy 366 cagaatcatgacctgaaagagagatgaatttcaagctgtgagccagggcagctcc 425
Db 194 CAGAAATCATGACCTGAAAGAGAGATGAAA-TTCAAGCTGTGAGCCAGGAGGCT-C 137
Qy 426 agtatggcaagggttcttgagaatcagccatttgggtacaaaaaagatttttaagcttt 485
Db 136 AGTATGGCAAGGCTTCTTGAGAAATCAGCCATTGTGTACAAAAGAGATTTTTAAAGCTTT 77
Qy 486 atgttataccatggagccatgagagagctatgattgttttaagactatttttaagcttt 545
Db 76 ATGTTATACCATGGAGCCATAGAAAGGCTATGATTTGTTAAGAACTATTTTAAAGTGT 17
Qy 546 ccagaccaccaaaaggga 561
Db 16 CCAGACCCCAAAAGGA 1

RESULT 14
AA626243/c
LOCUS
DEFINITION
IMAGE: zu93a08.s1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:745526
IMAGE: 3', mRNA sequence.
AA626243
ACCESSION
AA626243.1 GI:2538630
VERSION
AA626243.1
KEYWORDS
EST
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 410)
Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,
Krizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin
J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F., Theising, B.,
White, Y., Wyllie, T., Waterston, R. and Wilson, R.
WashU-NCI human EST Project
Unpublished (1997)
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.wustl.edu
This clone is available royalty-free through LLNL: contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 431.
Location/Qualifiers
1. 432
/organism="Homo sapiens"
/db_xref="GDB:1286258"
/db_xref="taxon:9606"
/clone="IMAGE:378002"
/clone_lib="Soares_fetal_heart_NDHH19W"
/sex="unknown"
FEATURES
Source

```

JOURNAL COMMENT	Unpublished (1997) Contact: Wilson RK Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: est@watson.wustl.edu This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information. Seq primer: -40m13 fwd. ET from Amersham High quality sequence stop: 370.
FEATURES	Location/Qualifiers 1..410 /organism="Homo sapiens" /db_xref="GDB:5933495" /db_xref="taxon:9606" /clone="IMAGE:745526" /sex="male" /lab_host="DH10B" /note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not 1; Site_2: Eco RI; 1st strand cDNA was prepared from mRNA obtained from Clontech Laboratories , Inc., and primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGAGCGCGCCCAATTTTITTTTITTTT 3']. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT713 vector. Library went through one round of normalization to Cot5, and was constructed by Bento Soares and M. Fatima Bonaldo.
BASE COUNT	116 a 89 c 81 g 124 t
ORIGIN	
Query Match	63.7%; Score 366; DB 9; Length 410;
Best Local Similarity	99.0%; Pred. No. 7.6e-63;
Matches 410; Conservative	0; Mismatches 0; Indels 4; Gaps 4;
QY	143 cttaagagggcagttctcaaaagcagaacatgccgcagctctcaagtttctctc 202
DB	410 CTTAAAGAGGGCAGTTTCTCAAAGCAGAAACATGCCGCCAGTTCTCAAGTTTTCCTCC 351
QY	203 taactccattgaatgtaagggcagctggcccccaatgtggaggtccgaacatttct 262
DB	350 TAATCCATTGTAATGTAAGGGCAGCTGGCCCCCAATGTGGGAGGTCCGAACATTTCT 291
QY	263 gaattccattttctgttcgcggctaataatgacagtttctgtctattcttagattcccca 322
DB	290 GAATTCCTCATTTCTTGTTCGGCGCTAAATGACAGTTTCTGTGATTACCTTAGATT-CCGA 232
QY	323 tctttcccaaggtgttgatttcaaaagggccagctaatagccagaaatcatgaccttg 382
DB	231 TCTTTCCCAAGGTTGTGATTACAAAGAGGCCACCTAATAG-CAGAAATCATGACCCCTG 173
QY	383 aaagagagatgaatttcaagctgtggccaggcagagctccagtgatggcaaggttct 442
DB	172 AAAGAGAGATCAAA-TTCAAGCTGTGAGCCAGCGCAGGAGCT-CAGTATGGCAAGGTTCT 115
QY	443 tgagaatcagccatttgggtacaaaagagatttttaagctttttatcccttgagc 502
DB	114 TGAGAAATCAGGCATTGTGTGTAACAAAAGAGATTTTAAAGCTTTTAAGTTTACCAAGGAGC 55
QY	503 catagaaggctatggattgttttaagaactatttttaagttgttccagacacaaa 556
DB	54 CATAGAAAGGCTATGGATTGTTTAAAGACTATTTTAAAGTGTTCAGACACAAA 1
RESULT 15	
A1341407/c	
LOCUS	q91d04.x1 NCI_CGAP_GC6 Homo sapiens cDNA clone IMAGE:2009863 3',
DEFINITION	mRNA sequence.
ACCESSION	A1341407
VERSION	A1341407.1 31:4078334
KEYWORDS	EST.
SOURCE	Human.
ORGANISM	Homo sapiens
REFERENCE	1 (bases 1 to 380) Mammalia; Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS	NCI-CGAP <a href="http://www.ncbi.nlm.nih.gov/ncicgap">http://www.ncbi.nlm.nih.gov/ncicgap</a> .
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP).
JOURNAL	Tumor Gene Index
COMMENT	Unpublished (1997) Contact: Robert Strausberg, Ph.D. Email: cgaps-r@mail.nih.gov Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D. cDNA Library Arrayed by: Greg Lennon, Ph.D. DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <a href="http://www-bio.llnl.gov/bbrp/image/image.html">www-bio.llnl.gov/bbrp/image/image.html</a> Insert Length: 628 Std Error: 0.00 Seq primer: -10UP from Gibco High quality sequence stop: 353.
FEATURES	Location/Qualifiers 1..380 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:2009863" /clone_lib="NCI_CGAP_GC6" /issue_type="pooled germ cell tumors" /lab_host="DH10B" /note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not 1; Site_2: Eco RI; Plasmid DNA from the normalized library NCI_CGAP_GC4 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (clonoids 1257096-1258631, 1469064-1470583, and 1475592-1476743). Subtraction by Bento Soares and M. Fatima Bonaldo.
BASE COUNT	110 a 82 c 76 g 112 t
ORIGIN	
Query Match	58.4%; Score 336; DB 9; Length 380;
Best Local Similarity	99.0%; Pred. No. 6.7e-57;
Matches 380; Conservative	0; Mismatches 0; Indels 4; Gaps 4;
QY	170 agaacatgccgcagttctcaagtcttctctcttaactccatttgaatglaagggcagct 229
DB	380 AGAACAATGCCGCCAGTCTCAAGTTTCTCTCTCACTCCATTTGAA1GTAAAGGCGAGCI 321
QY	230 ggcgcccaatgtgagggtccgaacatttctgaattccatttctgttccgcgctca 289
DB	320 GGCSCCAATGTGGGAGGTCCGAACATTTCTGAAATTCCTTTCTGTTCGCCGCTA 261
QY	250 aatgacagtttctgtcattacttacttccgcagctcttcccaagtggtgttttcaaaa 349
DB	260 AATGACAGTTTCTGTGATTACCTTAGATT-CCGATCTTCCCAAGGTGTGATTACAAA 202
QY	350 gaggcagctcaatgacagaaatcatgaccttgaagagagatgaatttcaagctgtga 409
DB	201 GAGGCCAGCTTAATAG-CAGAAATCATGACCTTGAAAGAGAGATGAAA-TTCAAGCTGTGA 144
QY	410 gccaggcaggcgtccagtgatggcaaggttcttgagaatcagccatttgggtacaaaaa 469
DB	143 GCCAGCGAGGAGCT-CAGTATGCAAGGTTCTTGAGAAATCAGCCATTTGTCACAAAAA 85
QY	470 gatttttaagcttttattgtatcaccatgagccatagaaaggctatggtattttaaaga 529
DB	84 GATTTTAAAGCTTTTAAGTTATACCATGAGCCATAGAAAGGCTATGGATTGTTTAAAGA 25

Tue Apr 30 14:18:16 2002

Qy 530 actattttaagtgttccagacc 553  
|||||  
Db 24 ACTATTTTANAAGTGTTCAGACC 1

Search completed: April 29, 2002, 17:17:22  
Job time: 13241 sec



polylinker; Site\_1: Not 1; Site\_2: Eco RI; 1st strand cDNA was prepared from human tonsillar cells enriched for germinal center B cells by flow sorting (CD20+, IgD-), provided by Dr. Louis M. Staudt (NCI), Dr. David Allman (NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was primed with a Not I - oligo(dT) primer [5'-TCTTACCAATCGAAGTGGAGCGCGCCCTCAATTTTTTTTTT-3' ]. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo.

BASE COUNT 163 a 62 c 60 g 155 t  
ORIGIN  
  
Query Match 72.5%; Score 428; DB 9; Length 440;  
Best Local Similarity 100.0%; Pred. No. 1.5e-51;  
Matches 428; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 1 gtgctctgtatttttttactgattgacattgaaatgaaatgcttgcgttaagatacgc 60  
Db 439 GTGCTCTGTGATTTTACTGATTAGACATTGATAGTAATTTGCGTTAAGATACGC 360  
  
Qy 61 ttaaggctcttctgaccatgttccctttgttagcaataaaatgttttttgcgaaact 120  
Db 379 TTAAGAGCTCTTTGTGACCATGTTCCCTTTGTAGCAATAAAATGTTTTTACGAAAC 320  
  
Qy 121 ttctccctggattgacagttttaaataaagacagatcattcaatgaaacagatattttaa 180  
Db 319 TTCTCCCTGGATTAGCAGTTTAAATGAACACAGAGTTTCATCAATGAAATGAGTAT 260  
  
Qy 181 taaaatttgccttaattgattcagttcagctcacaagtatttttaagatgattgagagac 240  
Db 259 TAAATATTTGCTTAATGATATCAGTTTCAAGCTCACAAGTATTTTAAAGATGAT 200  
  
Qy 241 ttgaattaaagaaaaaaatttcaatcatattttttaaatacaagataaatttttt 300  
Db 199 TTGAATTAAAGAAAAAAATTCATCATATTTTAAATATTAAGACTAAATTTGTT 140  
  
Qy 301 ttaaacacatttcaaatagagtgagtttgaactgacctatttttactcttttttaagt 360  
Db 139 TTAACAACACATTTCAATAGAAAGTGAGTTTGAACGTACCTTATTATACCTTTTAA 60  
  
Qy 361 ttgttcccttccctgtgcctgtgtcacaatttcaaatcttgcgtgaaatacatttgata 420  
Db 79 TTGTTCTCTTCCCTGTGCTGTGTCAAATCTTCAAGCTTTGCTGAAATACATTGATA 20  
  
Qy 421 caaagt 428  
Db 19 CAAAGTTT 12

RESULT 2  
AA857943/c  
LOCUS  
DEFINITION  
o169e08.s1 NCI\_CGAP\_C08 Homo sapiens cDNA clone IMAGE:1435622 3',  
mRNA sequence.  
ACCESSION  
AA857943  
VERSION  
AA857943.1 GI:2946245  
KEYWORDS  
EST.  
SOURCE  
human.  
ORGANISM  
Homo sapiens  
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1. (bases 1 to 440)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
Tumor Gene Index  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP).  
AUTHORS  
Unpublished (1997)  
JOURNAL  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
COMMENT  
\* Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.

Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation: M. Bento Soares, Ph.D.  
cDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LINL at:  
[www.bio.lnlni.gov/bbrp/image/image.html](http://www.bio.lnlni.gov/bbrp/image/image.html)  
Insert length: 646 Std Error: 0.00  
Seq primer: -10ml3 fwd, ET from Amersham  
High quality sequence stop: 380.  
Location/Qualifiers  
1. 440  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1435622"  
/clone\_lib="NCI\_CGAP\_C08"  
/tissue\_type="adenocarcinoma"  
/lab\_host="DH10B"  
/note="Organ: colon; Vector: pT7130-Pac (Pharmacia) with a  
modified polylinker; 1st strand cDNA was prepared from  
colon adenocarcinoma, and was then primed with a Not I -  
oligo(dT) primer. Double-stranded cDNA was ligated to Eco  
RI adaptors (Pharmacia), digested with Not I and cloned  
into the Not I and Eco RI sites of the modified pT73  
vector. Library is normalized. Library was constructed by  
Bento Soares and M. Fatima Bonaldo."

BASE COUNT 166 a 62 c 60 g 152 t  
ORIGIN  
  
Query Match 72.4%; Score 427; DB 9; Length 440;  
Best Local Similarity 100.0%; Pred. No. 2e-51;  
Matches 427; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 1 gtgctctgtatttttttactgattgacattgaaatgaaatgcttgcgttaagatacgc 60  
Db 434 GTGCTCTGTGATTTTACTGCAATGATAGTAATTTGCGTTAAGATACGC 375  
  
Qy 61 ttaaggctcttctgaccatgttccctttgttagcaataaaatgttttttgcgaaact 120  
Db 374 TTAAGGCTCTTTGTGACCATGTTTCCCTTTGAGCAATTAATGTTTTTACGAAAC 315  
  
Qy 121 ttctccctggattgacagttttaaataaagacagatcattcaatgaaatgagatattttaa 180  
Db 314 TTCTCCCTGGATTAGCAGTTTAAATGAACACAGAGTTTCATCAATGAAATGAGTAT 255  
  
Qy 181 ttaaaatttgcctttaaattgattcagttcagctcacaagtatttttaagatgattgagagac 240  
Db 254 TAAAAATTTGCCTTAATGATATCAGTTTCAAGCTCACAAGTATTTTAAATGAT 195  
  
Qy 241 ttgaattaaagaaaaaaatttctcaatcatattttttaaataaagataaattgttt 300  
Db 194 TTGAATTAAAGAAAAAAATTTCTCAATCATATTTTAAAAATAAGACTAAATTTG 135  
  
Qy 301 ttaaacacatttcaaatagagtgagtttgaactgacctatttttactcttttttaagt 360  
Db 134 TTAACAACATTTCAATAAGATGAGTTTGAAGTACCTTATTTTACTCTTTTAAAG 75  
  
Qy 361 ttgttcccttccctgtgcctgtgtcacaatttcaagcttctgcgtgaaatacatttgata 420  
Db 74 TTGTTCTCTTCCCTGTGCTGTGTCAAATCTTCAAGCTTTCGCTGAAATACATTGATA 15  
  
Qy 421 caaagt 427  
Db 14 CAAAGTT 8

RESULT 3  
BE047111/c  
LOCUS  
DEFINITION  
hg64903.x1 NCI\_CGAP\_HN13 Homo sapiens cDNA clone IMAGE:3124180 3',  
mRNA sequence.  
ACCESSION  
BE047111





367 TTAAGAGCTCTTTGTGACCAAGTTTCCCTTTGTAGCAATAAATGTTTTIACGAAACACI 308  
 121 tctccctggattagcagttttaaataagaacagagtttcacatgaatgagattattaa 180  
 307 TTTCTCCCTGGATTAGCAGTTTAAATGAACAGAGTTTCATCAATGAATGAGTAITATAAA 248  
 181 taaaaatttgcttaagtatcagttcagctcacaagctattttaaagatgattgagaagac 240  
 247 TAAAAATTGGCTTAATGTATCAGTTCAGCTCCCAAGTATTTTAAAGATGATGAGAAGAC 188  
 241 ttgaattaaag-aaaaaaaattctcaatcatattttttaaataataagaactaaattgtt 295  
 187 TTGAATTAAAGAAAAAAATTTCAATCATATTTTAAATATATAAGACTAAATTTGTT 128  
 300 tttaaacacatttcaaatagaatgagtttgaaactgaccttatttatactctttttaa 359  
 127 TTTAAACACATTTTCAATGAAGTGAAGTTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 68  
 360 ttgttcccttttccctgtgctgtgcaaatcttcaagctcttctgcaaaacacatttgat 419  
 67 TTTGTCTCTTTCCCTGGCTGTGTCAAATCTTCAAGTCTTGTGCTGAAATACATTTGAT 8  
 420 acaaaagt 426  
 7 CCAAGT 1

RESULT 7  
 AI690470/c 460 bp mRNA linear EST 16-DEC-1999  
 LOCUS tx98d12.x1 NCI\_CGAP\_Ut4 Homo sapiens cDNA clone IMAGE:2277599 3',  
 DEFINITION mRNA sequence.  
 ACCESSION AI690470  
 VERSION AI690470.1 GI:4901772  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo  
 1 (bases 1 to 460)  
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 Unpublished (1997)  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgapsb@femail.nih.gov  
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
 Emmert-Buck, M.D., Ph.D.  
 cDNA Library Preparation: Life Technologies, Inc.  
 DNA Sequencing by: Greg Lennon, Ph.D.  
 Cloning Distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 www-bio.llnl.gov/bbrp/image/image.html  
 Insert Length: 1322 Std Error: 0.00  
 Seq primer: -40UP from Gibco  
 High quality sequence stop: 414.  
 Location/Qualifiers  
 1..460  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:2277599"  
 /clone\_lib="NCI\_CGAP\_Ut4"  
 /tissue\_type="serous papillary carcinoma, high grade, 2  
 pooled tumors"  
 /lab\_host="DH108"  
 /note="Organ: uterus; Vector: pCMV-SPORT6; Site: 1; Salt:  
 Site 2: NotI; Cloned unidirectionally. Primer: Oligo di-  
 Average insert size 1.48 kb. Life Technologies catalog #:  
 11542-016"  
 171 a 69 c 65 g 155 t

BASE COUNT  
 ORIGIN

Query Match. 68.8%; Score 406; DB %; Length 460;  
 Best Local Similarity 99.5%; Pred. No. 1,8e-48;  
 Matches 428; Conservative 0; Mismatches 0; Indels 2; Gaps 2;  
 1 gtcctctgtgtatttttactgcattagacattgaatgaatgattgcttaagatacgc 60  
 430 GTGCTGTGTATTTTACTGCATTAGACATGTAATTTGCGTTAAGATACGC 371  
 61 ttaagggtctctgtgaccatgtttcccttctgagcaataaaatgttttttcgaaaaact 120  
 370 TTAAGGCTCTTTGTGACCATTGTCCTTTGTAGCAATAAATGTTTTTACGAAACT 311  
 121 tttcccttgattagcagttttaaataagaacagagtttcaatcaatgaatgagttattttaa 180  
 310 TTCTCCCTGGATTAGCAGTTTAAATGAACAGAGTTTCATCAATCAATGATGATGAGA 251  
 181 taaagaatttgcttcaatgattcagttcagttcagct-cacaagatttttcaagattgagaaga 239  
 250 TAAAAATTGGCTTAATGTATCAGTTCAGCTCCCAAGTATTTTAAATGATGATGAGA 191  
 240 ctgaattcaag-aaaaaaaattctcaatcatattttttaaataataagaactaaattgt 298  
 190 CTGAATTAAAGAAAAAAATTTCTCAATCATATTTTAAATATATAAGACTAAAAATTGT 131  
 259 tttttaaacaacatttcaaatagaatgagtttgaaactgaccttatttatactcttttaa 358  
 130 TTTAAACACATTTTCAATGAAGTGAAGTTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 71  
 359 gtttcttcttcttctgctgtgctgtgcaaatcttcaagctcttctgcaaaacacatttga 418  
 70 GTTGTCTCTTTCCCTGGCTGTGTCAAATCTTCAAGTCTTGTGCAAAATACATTGA 11  
 419 tacaagaattt 428  
 10 TACAAAGTTT 1

RESULT 8  
 AA258236/c 411 bp mRNA linear EST 13-AUG-1997  
 LOCUS 2533d04.x1 NCI\_CGAP\_GCBI Homo sapiens cDNA clone IMAGE:686983 3',  
 DEFINITION mRNA sequence.  
 ACCESSION AA258236  
 VERSION AA258236.1 GI:1894231  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 1 (bases 1 to 411)  
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 Unpublished (1997)  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgapsb@femail.nih.gov  
 This clone is available royalty-free through LLNL; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 Insert Length: 953 Std Error: 0.00  
 Seq primer: -41ml3 fwd. ET from Amersham  
 High quality sequence stop: 324.  
 Location/Qualifiers  
 1..411  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:686983"  
 /clone\_lib="NCI\_CGAP\_GCBI"  
 /tissue\_type="germinal center B cell"  
 /lab\_host="DH108"  
 /note="Vector: pT7T3D-Pac (Pharmacia) with a modified  
 polylinker; Site: 1; Not I; Site: 2; Eco RI; 1st strand cDNA  
 was prepared from human tonsillar cells enriched for





CONTACT: Robert Strausberg, Ph.D.  
Email: rcsap@fmail.nih.gov  
Tissue Procurement: Chris Moskaluk, M.D., Ph.D., Michael R. Emert-Buck, M.D., Ph.D.  
cDNA Library Preparation: Life Technologies, Inc.  
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LINL)  
Sequence Generated by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LINL, send email to: info@image.lnl.gov

High quality sequence stop: 404.  
Location/Qualifiers  
1. .410  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3653518"  
/clone\_lib="NCI\_CGAP\_Skn1"  
/lab\_host="DH10B (T1 phage-resistant)"  
/note="Organ: skin, normal, 4 pooled samples; Vector: PCMV-SpOtk6; Site\_1: NotI; Site\_2: SalI; Cloned unidirectionally. Primer: oligo dr. Average insert size 1.03kb. Library constructed by Life Technologies. cat # 11132-018."

148 a 58 c 57 g 147 t

BASE COUNT 148 a 58 c 57 g 147 t

ORIGIN

Query Match 66.4%; Score 391.8; DB 10; Length 410;  
Best Local Similarity 99.3%; Pred. NO. 1.9e-46;  
Matches 404; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 23 gcattagacattgaatgaatttcggttaagatagcgttcaaaagctctttgtgaccatg 82  
|||||  
DB 410 GCATTAGACATGAATAGTAATTCGCTTAAGATACGCTTAAAGGCTCTTTGTGACCATT 351  
|||||

QY 83 ttccctttgtgacacataaagtgttttttttacgaaacattcttccctgattgagcttta 142  
|||||

DB 350 TTTCCTCTTGTAGCAATAAAATGTTTTTACGAAACATTTCCCTGGATTAGCAGHTTA 291  
|||||

QY 143 aatgaacagaggtctcatcgaatgaatgagttatataaaataaaatttccttaattgtatc 202  
|||||

DB 250 AATGAACAGAGGTTTCATCAATCAATGAATGAGTATTTAAATAAAATTTGCTTTAATCTAIC 231  
|||||

QY 203 agtcagctcacaagtatttttaagatgattgagaagactgaattcaag-aaaaaataat 261  
|||||

DB 230 AGTTCAGCTCACAAAGTATTTTAAAGATGATTGAGACACTTGAATTAAGAAATAAAAT 171  
|||||

QY 262 tctcaatcatattttlaaataataagactcaaaattgttttttlaaaacacattttcaaataga 321  
|||||

DB 170 TCTCAATCATATTTTAAATAATAAGACTAAATTTGTTTTTAAACACATTTTCAAAATAGA 111  
|||||

QY 322 agtgagttgaactgacctattttactctttttaaagtttctcttcttccctgtgctt 381  
|||||

DB 110 AGTGAGTTGAACGTGACCTATTATTATCTCTTTTAAAGTTTGTCTCTTTCCTGTGCT 51  
|||||

QY 382 gtgtcaaatcttcaagtcttgcgtgaataacacattttgatgatacaagttt 428  
|||||

DB 50 GTGTCAAACTTCAAGCTTTCGTGCTGAAATACATTTGATACAAAGTTT 4  
|||||

RESULT 11  
AA150563/c  
LOCUS  
DEFINITION 418 bp mRNA linear EST 15-MAY-1997  
IMAGE:504864 3', mRNA sequence.  
AA150563  
AA150563.1 GI:1722474  
EST.  
human.  
SOURCE  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
1 (bases 1 to 418)

REFERENCE 1

CONTACT: Robert Strausberg, Ph.D.  
Email: rcsap@fmail.nih.gov  
Tissue Procurement: Chris Moskaluk, M.D., Ph.D., Michael R. Emert-Buck, M.D., Ph.D.  
cDNA Library Preparation: Life Technologies, Inc.  
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LINL)  
Sequence Generated by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LINL, send email to: info@image.lnl.gov

High quality sequence stop: 404.  
Location/Qualifiers  
1. .410  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3653518"  
/clone\_lib="NCI\_CGAP\_Skn1"  
/lab\_host="DH10B (T1 phage-resistant)"  
/note="Organ: skin, normal, 4 pooled samples; Vector: PCMV-SpOtk6; Site\_1: NotI; Site\_2: SalI; Cloned unidirectionally. Primer: oligo dr. Average insert size 1.03kb. Library constructed by Life Technologies. cat # 11132-018."

148 a 58 c 57 g 147 t

BASE COUNT 148 a 58 c 57 g 147 t

ORIGIN

Query Match 66.4%; Score 391.8; DB 10; Length 410;  
Best Local Similarity 99.3%; Pred. NO. 1.9e-46;  
Matches 404; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 23 gcattagacattgaatgaatttcggttaagatagcgttcaaaagctctttgtgaccatg 82  
|||||  
DB 410 GCATTAGACATGAATAGTAATTCGCTTAAGATACGCTTAAAGGCTCTTTGTGACCATT 351  
|||||

QY 83 ttccctttgtgacacataaagtgttttttttacgaaacattcttccctgattgagcttta 142  
|||||

DB 350 TTTCCTCTTGTAGCAATAAAATGTTTTTACGAAACATTTCCCTGGATTAGCAGHTTA 291  
|||||

QY 143 aatgaacagaggtctcatcgaatgaatgagttatataaaataaaatttccttaattgtatc 202  
|||||

DB 250 AATGAACAGAGGTTTCATCAATCAATGAATGAGTATTTAAATAAAATTTGCTTTAATCTAIC 231  
|||||

QY 203 agtcagctcacaagtatttttaagatgattgagaagactgaattcaag-aaaaaataat 261  
|||||

DB 230 AGTTCAGCTCACAAAGTATTTTAAAGATGATTGAGACACTTGAATTAAGAAATAAAAT 171  
|||||

QY 262 tctcaatcatattttlaaataataagactcaaaattgttttttlaaaacacattttcaaataga 321  
|||||

DB 170 TCTCAATCATATTTTAAATAATAAGACTAAATTTGTTTTTAAACACATTTTCAAAATAGA 111  
|||||

QY 322 agtgagttgaactgacctattttactctttttaaagtttctcttcttccctgtgctt 381  
|||||

DB 110 AGTGAGTTGAACGTGACCTATTATTATCTCTTTTAAAGTTTGTCTCTTTCCTGTGCT 51  
|||||

QY 382 gtgtcaaatcttcaagtcttgcgtgaataacacattttgatgatacaagttt 428  
|||||

DB 50 GTGTCAAACTTCAAGCTTTCGTGCTGAAATACATTTGATACAAAGTTT 4  
|||||

RESULT 11  
AA150563/c  
LOCUS  
DEFINITION 418 bp mRNA linear EST 15-MAY-1997  
IMAGE:504864 3', mRNA sequence.  
AA150563  
AA150563.1 GI:1722474  
EST.  
human.  
SOURCE  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
1 (bases 1 to 418)

REFERENCE 1



with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTAACATCTGAGGAGGCGCCGCTTTTCTTTTCTTTT 3'] , double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT73 vector (Pharmacia). Library constructed by Bento Soares and M. Fatima Bonaldo.

BASE COUNT 129 a 46 c 50 g 135 t  
ORIGIN

Query Match 57.6%; Score 340; DB 9; Length 360;  
Best Local Similarity 99.7%; Pred. No. 4.2e-39;  
Matches 351; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 77 accatgtttccctttgttagcaataaaatgtttttacgaaacatttccctggattagc 136  
DB 360 ACCATGTTCCCTTTGTAGCAATAAAATGTTTTCAGAAACATTCCTCGGATTAGC 301  
QY 137 agtttaaaagaaacagagtttcatcaatgaaatgagttatttaaaataaaatgttcttaa 196  
DB 360 AGTTTAAATCAACACAGAGTTCATCAATGAAATGAGTATTTAAATTAATAATTTGCTTAA 241  
QY 197 tgtatcagttcagctcacagatttttaagatgattgagaagactgaattaaag-aaa 255  
DB 240 TGTATCAGTTCAGCTCACAGATTATTTAAGATGATTGAGAAGACTTCAATTAAGAAAAA 181  
QY 256 aaaaattctcaatcatatttttaaaataaagactaaaatgtttttaaacaacatttca 315  
DB 180 AAAAAATTCATATATTTTAAATAAAGACTAAATATGTTTTTAAACACATTTCA 121  
QY 316 aatagaagtgtgtgaactgacctattttatactcttttaagtgttcttcttccct 375  
DB 120 AATAGAAGTCAGTTGAAGTCACCTTATTATATCTCTTTTAAAGTTGTCTCTTTCCCT 61  
QY 376 gtgcctgtgcaaatcttcaagctctgtgcaaaaatacatattgatacaagtt 427  
DB 60 GTGCCTGTGCAAAATCTTCAAGTCTTCTGCAAAATACATTTGATACAAAGTT 9

RESULT 15  
LOCUS AW023325  
DEFINITION IMAGE:2487136 5', mRNA sequence.  
ACCESSION AW023325  
VERSION AW023325.1 GI:5876855  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 467)  
AUTHORS Robertson,N.G., Khetarpal,U., Gutierrez-Espeleta,G.A., Bleber,F.R. and Morton,C.C.  
TITLE Isolation of novel and known genes from a human fetal cochlear cDNA library using subtractive hybridization and differential screening  
JOURNAL Genomics 23, 42-50 (1994)  
MEDLINE 9513011  
COMMENT Contact: Morton, C. C. Departments of Pathology and Obstetrics, Gynecology and Reproductive Biology Brigham and Women's Hospital 75 Francis Street, Harvard Medical School, Boston, MA 02115, USA Tel: 617 732 7980 Fax: 617 738 6996 Email: cmorton@rics.bwh.harvard.edu DNA sequencing and analyses were performed by National Institutes of Health Intramural Sequencing Center (NISC; see http://www.nisc.nih.gov). This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

Query Match 59.3%; Score 350; DB 9; Length 370;  
Best Local Similarity 99.7%; Pred. No. 1.6e-40;  
Matches 361; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 67 gctctttgacacatttccctttgttagcaataaaatgtttttacgaaacatttctcc 126  
DB 370 GCTCTTTGTGACCATGTTCCCTTTGTAGCAATAAAATGTTTTCAGAAACATTTCTTCC 311  
QY 127 ctggattagcagtttaaatgaaacagagtttcatcaatgaaatgagttattttaaataaaaa 186  
DB 310 CTGGATTAGCAGTTTAAATGAAACAGAGTTCATCAATGAAATGAGTATTTTAAATAAAAA 251  
QY 187 ttgctcttaagtatcagttcagctcacagatttttaagatgattgagaagacttgaat 246  
DB 250 TTGCTCTTAAGTATCAGTTTCAAGTTCAGCTCACAGTATTTTAAAGATGATTGAGAAGAC 191  
QY 247 taaag-aaaaaaattctcaatcatatttttaaaataaataagactaaaattgtttttaa 305  
DB 190 TAAAGAAAAAATAATTTCTCAATCATATTTTAAATATAAGACTAAAAATGTTTAAAA 131  
QY 306 acacatttcaaatagaagttgttgaactgacctattttatcacctcttttcaagtttgtt 365  
DB 130 ACACATTTCAAATAGAAGTGTGTTGAAGTACCTTATTTATATCTCTTTTAAAGTTGT 71  
QY 366 cctttccctgtgctgtgcaaatcttcaagctctgtgcaaaaatacatatttgatacaaa 425  
DB 70 CCTTTCCCTGTGCTGTGCTCAAAATCTTCAAGTCTTCTGCAAAATACATTTGATACAAAG 11  
QY 426 tt 427  
DB 10 TT 9

RESULT 14  
LOCUS AA701126/c  
DEFINITION IMAGE:397360 3', mRNA sequence.  
ACCESSION AA701126  
VERSION AA701126.1 GI:2704291  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 360)  
AUTHORS Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisli,G., Jost,S., Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.  
TITLE WashU-NCI human EST Project  
JOURNAL Unpublished (1997)  
COMMENT Contact: Wilson RK Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: est@wustl.edu This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information. Seq primer: -40ml3 fwd. ET from Amersham High quality sequence stop: 322.  
FEATURES Location/Qualifiers  
1..360  
/organism="Homo sapiens"  
/db\_xref="GDB:1303262"  
/db\_xref="taxon:9606"  
/clone="IMAGE:397360"  
/clone\_lib="Soares\_pineal\_gland\_N3HPG"  
/lab\_host="DH10B (ampicillin resistant)"  
/note="Organ: pineal gland; Vector: pT713D (Pharmacia)

Tue Apr 30 14:18:20 2002

us-09-248-178-67.std.rst

```
Plate: ILAM6189 row: G column: 5
Seq primer: M13RPI reverse primer (ABI).
FEATURES
    Location/Qualifiers
        1..467
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="IMAGE:2487196"
            /clone_lib="Morton Fetal Cochlea"
            /tissue_type="cochlea"
            /dev_stage="16-22 week fetus"
            /lab_host="SOLR cells (kanamycin resistant)"
            /note="Organ: ear; Vector: pBluescript SK-; Site: 1: ECORI;
            Site_2: XhoI; Reference: Genomics 23, 42-50 (1994) Cloned
            unidirectionally. Primer: Oligo dT. Fetal cochlea, normal.
            37% of inserts <0.5 kb, 56% 0.5-1.0 kb, 7% >1 kb. Uni-ZAP
            XR Vector. Library constructed by N. Robertson, C. Morton.
            -5' adaptor sequence: 5' GAATTCGCCACGAG 3' -3' adaptor
            sequence: 5' CTCGAGTTTCTTTTCTTTT 3'.
```

BASE COUNT	161 a	62 c	74 g	170 t
ORIGIN				
Query Match	54.1%	Score 319;	DB 9;	Length 467;
Best Local Similarity	99.4%	Pred. No. 3.5e-36;		
Matches 341;	Conservative 0;	Mismatches 0;	Indels 2;	Gaps 2;
QY	1	gtgctctgtgtatttttttactgcatagacattgaatagtaatttgcgttaagatcacgc	60	
Db	125	GTGCTCTGTGTATTTTTTACTGTCATTAGACATTCAATAGTAGTAATTCGGTTAAGATACGC	184	
QY	61	ttaaaggctcttctgtgaccatgtttcccttctgtagcaataaaatgttttttaccgaaaact	120	
Db	185	TTAAAGGCTCTTCTGACCATGTTCCCTTTGTAGCAATAAAATGTTTTTTACGAAACI	244	
QY	121	ttctccctgattagcagtttaaatgaaacagagttcatcaatgaaatgagtgatttaaaa	180	
Db	215	TTCTCCCTGGATTAGCAGTTTAAATCAACACAGAGTTCATCAATGAATGAGTATTAAAA	304	
QY	181	taaaaatttgcttaattgtatc-agttcaggtcacaaagtatttttaagatgattgagaaga	239	
Db	305	TAAAAATTTCCTTAAATGTATCAAGTTCAGCTCACAGTATTTTAAAGATGATTGAGAAGA	364	
QY	240	cttgaatttaagaaaaaaattctcaatcataatttttaaaatat-aagactaaaaattgt	298	
Db	365	CTTGAAATTAAAGAAAAAAAATTCCTCAATCATATTTTAAATATAAAGACTAAAAATTGT	424	
QY	299	ttttaaacacatttcaaatagagtgagtttgaaactgacctt	341	
Db	425	TTTTAAACACATTTCAATAGAGAGTGAGTTTGAACCTGACCTT	467	

Search completed: April 29, 2002, 17:17:25  
Job time: 13244 sec



Nature Biotechnology, in press. Note that even though the cell type indicated is H1080, since a random activation method was used, these sequence tags are not necessarily expressed in H1080 under normal circumstances."

BASE COUNT 222 a 168 c 139 g 208 t  
ORIGIN

Query Match 32.3% Score 301; DB 11; Length 737;  
Best Local Similarity 100.0%; Pred. No. 8.1e-83;  
Matches 301; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 taccgggttactcgacaaaccacccatcccttccagatgagcgccttaccctgcagaa 224  
DB 77 TATCCCGGTACTCGACAAACCCATCACCTTCCAGATGAGCGCTTACCTGCAGGAA 136  
QY 225 taactgttttcaaatatttgggtcttcacacacacccctatttcttgggaagaccctc 284  
DB 137 TAACGTGTTTCAATATTGGGCTTTCACCAACCCCTATTTCIGGAAGACCCCTC 196  
QY 285 aggtctttaaccccttgagattctccagggaataattctgaaaaaacatccctatgcct 344  
DB 197 AGGCTTTAAACCCCTTCCAGGGAATAATCTGAAAAAATACATCCCTATGCCT 256  
QY 345 tcataccattctcagctgagattaaggaactgcattgggcagcatttgcataaattgagt 404  
DB 257 TCATACCATTCTCAGCTGGATTAGGAACCTCATTTGGCAGCATTTTGCATTAATGAGT 316  
QY 405 gtaagtggcagtggttaactctgcctgccttcaagctggctccagaccaccctcaagc 464  
DB 317 GTAAAGTGGCAGTGCCATTAACTCTGCTCGCTTCAAGCTGGCTCCAGACCACTCAAGC 376  
QY 465 c 465  
DB 377 c 377

RESULT 2  
BG101072 788 bp mRNA EST 21-APR-2001  
LOCUS  
DEFINITION RST20398 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.  
ACCESSION BG201072  
VERSION BG201072.1 GI:13722759  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 788)  
AUTHORS Harrington,J.J., Sherf,B., Rundlett,S., Jackson,P.D., Perry,R., Cain,S., Dahl,T., Thornton,M., Ramachandran,R., Whittington,J., Lerner,L., Krashoc,D., McElligott,K., Clark,S., Mays,R., Smith,E., Veloso,N., Hess,J., Cothren,K., Lo,K., Offenbacher,J., Danzig,J. and Ducar,M.  
TITLE Creation of Genome-wide Protein Expression Libraries using Random Activation of Gene Expression  
JOURNAL Nat. Biotechnol. 19 (5), 440 (2001) In press  
COMMENT Contact: Scott J. Cain  
Athersys, Inc.  
3201 Carnegie Ave, Cleveland, OH 44115, USA  
Tel: 216 431 9900  
Fax: 216 361 9596  
Email: scai@atersys.com  
High quality sequence stop: 514.  
Location/Qualifiers  
1. .788  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="Athersys RAGE Library"  
/cell\_line="H1080"  
/note="See 'Creation of Genome-wide Protein Expression Libraries using Random Activation of Gene Expression', Nature Biotechnology, in press. Note that even though the

cell type indicated is H1080, since a random activation method was used, these sequence tags are not necessarily expressed in H1080 under normal circumstances."

BASE COUNT 236 a 141 c 146 g 221 t  
ORIGIN

Query Match 32.3% Score 301; DB 11; Length 788;  
Best Local Similarity 100.0%; Pred. No. 7.6e-83;  
Matches 301; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 taccgggttactcgacaaaccacccatcccttccagatgagcgccttaccctgcagaa 224  
DB 77 TATCCCGGTACTCGACAAACCCATCACCTTCCAGATGAGCGCTTACCTGCAGGAA 136  
QY 225 taactgttttcaaatatttgggtcttcacacacacccctatttcttgggaagaccctc 284  
DB 137 TAACGTGTTTCAATATTGGGCTTTCACCAACCCCTATTTCIGGAAGACCCCTC 196  
QY 285 aggtctttaaccccttgagattctccagggaataattctgaaaaaacatccctatgcct 344  
DB 197 AGGCTTTAAACCCCTTCCAGGGAATAATCTGAAAAAATACATCCCTATGCCT 256  
QY 345 tcataccattctcagctgagattaaggaactgcattgggcagcatttgcataaattgagt 404  
DB 257 TCATACCATTCTCAGCTGGATTAGGAACCTCATTTGGCAGCATTTTGCATTAATGAGT 316  
QY 405 gtaagtggcagtggttaactctgcctgccttcaagctggctccagaccaccctcaagc 464  
DB 317 GTAAAGTGGCAGTGCCATTAACTCTGCTCGCTTCAAGCTGGCTCCAGACCACTCAAGC 376  
QY 465 c 465  
DB 377 c 377

RESULT 3  
BG193485 792 bp mRNA EST 21-APR-2001  
LOCUS  
DEFINITION RST12619 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.  
ACCESSION BG193485  
VERSION BG193485.1 GI:13715172  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 792)  
AUTHORS Harrington,J.J., Sherf,B., Rundlett,S., Jackson,P.D., Perry,R., Cain,S., Dahl,T., Thornton,M., Ramachandran,R., Whittington,J., Lerner,L., Krashoc,D., McElligott,K., Clark,S., Mays,R., Smith,E., Veloso,N., Hess,J., Cothren,K., Lo,K., Offenbacher,J., Danzig,J. and Ducar,M.  
TITLE Creation of Genome-wide Protein Expression Libraries using Random Activation of Gene Expression  
JOURNAL Nat. Biotechnol. 19 (5), 440 (2001) In press  
COMMENT Contact: Scott J. Cain  
Athersys, Inc.  
3201 Carnegie Ave, Cleveland, OH 44115, USA  
Tel: 216 431 9900  
Fax: 216 361 9596  
Email: scai@atersys.com  
High quality sequence stop: 553.  
Location/Qualifiers  
1. .792  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="Athersys RAGE Library"  
/cell\_line="H1080"  
/note="See 'Creation of Genome-wide Protein Expression Libraries using Random Activation of Gene Expression', Nature Biotechnology, in press. Note that even though the

Tue Apr 30 14:18:38 2002

```

method was used, these sequence tags are not necessarily
expressed in H1080 under normal circumstances.
BASE COUNT      222 a 190 c 158 g 222 t
ORIGIN
Query Match      31.5%; Score 294; DB 11; Length 792;
Best Local Similarity 99.6%; Pred. No. 1e-80; 1: Indels 1: Gaps 1:
Matches 464; Conservative 0; Mismatches 1:
QY 1 tttactgttgcgaatgacctgagcatcagcagagatgccgagatgaatcagggaac 60
DB 108 tttactgttgcgaatgacctgagcatcagcagagatgccgagatgaatcagggaac 167
QY 61 tcttagggatgggtcttctattaccctgggaacacctgagcagatgccctaccacaga 120
DB 168 tcttagggatgggtcttctattaccctgggaacacctgagcagatgccctaccacaga 227
QY 121 tggatcaagaatgacctcgcctctacgcacggtagtaaac-tatccggttactcg 179
DB 228 tggatcaagaatgacctcgcctctacgcacggtagtaaac-tatccggttactcg 287
QY 180 acaaacccatccattccagatggagcgtctctaccctgagcagaataactgtgtttatca 239
DB 288 acaaacccatccattccagatggagcgtctctaccctgagcagaataactgtgtttatca 347
QY 240 atattggcctcttcacacacacctctattctgggaagaccctcaggtgttttaacccct 299
DB 348 atattggcctcttcacacacacctctattctgggaagaccctcaggtgttttaacccct 407
QY 300 tggattctccagggaatattctgaaaaatacatccctatgcctcattaccattctcag 359
DB 408 tggattctccagggaatattctgaaaaatacatccctatgcctcattaccattctcag 467
QY 360 ctggattaagaactcattgggcagcattttggccataattgagtgtaagtggcagtg 419
DB 468 ctggattaagaactcattgggcagcattttggccataattgagtgtaagtggcagtg 527
QY 420 cattaactctgcctcgtcttaagctggctccagacacctcaagggc 465
DB 528 cattaactctgcctcgtcttaagctggctccagacacctcaagggc 573
RESULT 4
BG219391 777 bp mRNA EST 21-APR-2001
LOCUS BG219391 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.
DEFINITION BG219391
ACCESSION BG219391
VERSION BG219391.1 GI:13745412
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Harrington,J.J., Sherf,B., Rundlett,S., Jackson,P.D., Perry,R.,
Cain,S., Dahl,T., Thornton,M., Ramachandran,R., Whittington,J.,
Lerner,L., Krashoc,D., McElligott,K., Clark,S., Mays,R., Smith,E.,
Veloso,N., Hess,J., Cothren,K., Lo,K., Offenbacher,J., Danzig,J.,
and Ducar,M.
Creation of Genome-wide Protein Expression Libraries using Random
Activation of Gene Expression
Nat. Biotechnol. 19 (5), 440 (2001) In press
Contact: Scott J. Cain
Athersys, Inc.
3201 Carnegie Ave, Cleveland, OH 44115, USA
Tel: 216 431 9900
Fax: 216 361 9596
Email: scain@athersys.com
High quality sequence stop: 529.
Location/Qualifiers
FEATURES
source
1. .777
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Athersys RAGE Library"
/cell_line="H1080"
/note="See 'Creation of Genome-wide Protein Expression
Libraries using Random Activation of Gene Expression',
method was used, these sequence tags are not necessarily
expressed in H1080 under normal circumstances."
BASE COUNT      222 a 190 c 158 g 222 t
ORIGIN
Query Match      31.5%; Score 294; DB 11; Length 792;
Best Local Similarity 99.6%; Pred. No. 1e-80; 1: Indels 1: Gaps 1:
Matches 464; Conservative 0; Mismatches 1:
QY 1 tttactgttgcgaatgacctgagcatcagcagagatgccgagatgaatcagggaac 60
DB 108 tttactgttgcgaatgacctgagcatcagcagagatgccgagatgaatcagggaac 167
QY 61 tcttagggatgggtcttctattaccctgggaacacctgagcagatgccctaccacaga 120
DB 168 tcttagggatgggtcttctattaccctgggaacacctgagcagatgccctaccacaga 227
QY 121 tggatcaagaatgacctcgcctctacgcacggtagtaaac-tatccggttactcg 179
DB 228 tggatcaagaatgacctcgcctctacgcacggtagtaaac-tatccggttactcg 287
QY 180 acaaacccatccattccagatggagcgtctctaccctgagcagaataactgtgtttatca 239
DB 288 acaaacccatccattccagatggagcgtctctaccctgagcagaataactgtgtttatca 347
QY 240 atattggcctcttcacacacacctctattctgggaagaccctcaggtgttttaacccct 299
DB 348 atattggcctcttcacacacacctctattctgggaagaccctcaggtgttttaacccct 407
QY 300 tggattctccagggaatattctgaaaaatacatccctatgcctcattaccattctcag 359
DB 408 tggattctccagggaatattctgaaaaatacatccctatgcctcattaccattctcag 467
QY 360 ctggattaagaactcattgggcagcattttggccataattgagtgtaagtggcagtg 419
DB 468 ctggattaagaactcattgggcagcattttggccataattgagtgtaagtggcagtg 527
QY 420 cattaactctgcctcgtcttaagctggctccagacacctcaagggc 465
DB 528 cattaactctgcctcgtcttaagctggctccagacacctcaagggc 573
RESULT 4
BG219391 777 bp mRNA EST 21-APR-2001
LOCUS BG219391 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.
DEFINITION BG219391
ACCESSION BG219391
VERSION BG219391.1 GI:13745412
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Harrington,J.J., Sherf,B., Rundlett,S., Jackson,P.D., Perry,R.,
Cain,S., Dahl,T., Thornton,M., Ramachandran,R., Whittington,J.,
Lerner,L., Krashoc,D., McElligott,K., Clark,S., Mays,R., Smith,E.,
Veloso,N., Hess,J., Cothren,K., Lo,K., Offenbacher,J., Danzig,J.,
and Ducar,M.
Creation of Genome-wide Protein Expression Libraries using Random
Activation of Gene Expression
Nat. Biotechnol. 19 (5), 440 (2001) In press
Contact: Scott J. Cain
Athersys, Inc.
3201 Carnegie Ave, Cleveland, OH 44115, USA
Tel: 216 431 9900
Fax: 216 361 9596
Email: scain@athersys.com
High quality sequence stop: 529.
Location/Qualifiers
FEATURES
source
1. .777
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Athersys RAGE Library"
/cell_line="H1080"
/note="See 'Creation of Genome-wide Protein Expression
Libraries using Random Activation of Gene Expression',
method was used, these sequence tags are not necessarily
expressed in H1080 under normal circumstances."
BASE COUNT      222 a 190 c 158 g 222 t
ORIGIN
Query Match      28.8%; Score 269; DB 11; Length 777;
Best Local Similarity 100.0%; Pred. No. 4.6e-73;
Matches 269; Conservative 0; Mismatches 0; Indels 0; Gaps 0:
QY 157 ccagatggagcgtccttaccctgcagggaataactgctttatcaatttggctcttcac 256
DB 111 ccagatggagcgtccttaccctgcagggaataactgctttatcaatttggctcttcac 170
QY 257 cacaacccctattctcgggaagaccctcaggtcttttaaccccttgagattctccaggaa 316
DB 171 cacaacccctattctcgggaagaccctcaggtcttttaaccccttgagattctccaggaa 230
QY 317 aattctgaaaaatacatccctatgcctcattaccattcagctgattcaaggaactgc 376
DB 231 aattctgaaaaatacatccctatgcctcattaccattcagctgattcaaggaactgc 290
QY 377 attgggcagcatttggccataattgagtgtaagtggaagtggtggtggtggtggtgc 436
DB 291 attgggcagcatttggccataattgagtgtaagtggaagtggtggtggtggtggtgc 350
QY 437 ttcaagctggctccagacacctcaagggc 465
DB 351 ttcaagctggctccagacacctcaagggc 379
RESULT 5
BG218514 787 bp mRNA EST 21-APR-2001
LOCUS BG218514 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.
DEFINITION BG218514
ACCESSION BG218514
VERSION BG218514.1 GI:13744423
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Harrington,J.J., Sherf,B., Rundlett,S., Jackson,P.D., Perry,R.,
Cain,S., Dahl,T., Thornton,M., Ramachandran,R., Whittington,J.,
Lerner,L., Krashoc,D., McElligott,K., Clark,S., Mays,R., Smith,E.,
Veloso,N., Hess,J., Cothren,K., Lo,K., Offenbacher,J., Danzig,J.,
and Ducar,M.
Creation of Genome-wide Protein Expression Libraries using Random
Activation of Gene Expression
Nat. Biotechnol. 19 (5), 440 (2001) In press
Contact: Scott J. Cain
Athersys, Inc.
3201 Carnegie Ave, Cleveland, OH 44115, USA
Tel: 216 431 9900
Fax: 216 361 9596
Email: scain@athersys.com
High quality sequence stop: 504.
Location/Qualifiers
FEATURES
source
1. .787
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Athersys RAGE Library"
/cell_line="H1080"
/note="See 'Creation of Genome-wide Protein Expression
Libraries using Random Activation of Gene Expression',
method was used, these sequence tags are not necessarily
expressed in H1080 under normal circumstances."
BASE COUNT      230 a 177 c 142 g 226 t
ORIGIN

```

```

/ab_xref="taxon:9606"
/clone_lib="Athersys RAGE Library"
/cell_line="H1080"
/note="See 'Creation of Genome-wide Protein Expression',
Libraries using Random Activation of Gene Expression',
Nature Biotechnology, in press. Note that even though the
cell type indicated is H1080, since a random activation
method was used, these sequence tags are not necessarily
expressed in H1080 under normal circumstances."
BASE COUNT      230 a 177 c 142 g 226 t
ORIGIN
Query Match      28.8%; Score 269; DB 11; Length 777;
Best Local Similarity 100.0%; Pred. No. 4.6e-73;
Matches 269; Conservative 0; Mismatches 0; Indels 0; Gaps 0:
QY 157 ccagatggagcgtccttaccctgcagggaataactgctttatcaatttggctcttcac 256
DB 111 ccagatggagcgtccttaccctgcagggaataactgctttatcaatttggctcttcac 170
QY 257 cacaacccctattctcgggaagaccctcaggtcttttaaccccttgagattctccaggaa 316
DB 171 cacaacccctattctcgggaagaccctcaggtcttttaaccccttgagattctccaggaa 230
QY 317 aattctgaaaaatacatccctatgcctcattaccattcagctgattcaaggaactgc 376
DB 231 aattctgaaaaatacatccctatgcctcattaccattcagctgattcaaggaactgc 290
QY 377 attgggcagcatttggccataattgagtgtaagtggaagtggtggtggtggtggtgc 436
DB 291 attgggcagcatttggccataattgagtgtaagtggaagtggtggtggtggtggtgc 350
QY 437 ttcaagctggctccagacacctcaagggc 465
DB 351 ttcaagctggctccagacacctcaagggc 379
RESULT 5
BG218514 787 bp mRNA EST 21-APR-2001
LOCUS BG218514 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.
DEFINITION BG218514
ACCESSION BG218514
VERSION BG218514.1 GI:13744423
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Harrington,J.J., Sherf,B., Rundlett,S., Jackson,P.D., Perry,R.,
Cain,S., Dahl,T., Thornton,M., Ramachandran,R., Whittington,J.,
Lerner,L., Krashoc,D., McElligott,K., Clark,S., Mays,R., Smith,E.,
Veloso,N., Hess,J., Cothren,K., Lo,K., Offenbacher,J., Danzig,J.,
and Ducar,M.
Creation of Genome-wide Protein Expression Libraries using Random
Activation of Gene Expression
Nat. Biotechnol. 19 (5), 440 (2001) In press
Contact: Scott J. Cain
Athersys, Inc.
3201 Carnegie Ave, Cleveland, OH 44115, USA
Tel: 216 431 9900
Fax: 216 361 9596
Email: scain@athersys.com
High quality sequence stop: 504.
Location/Qualifiers
FEATURES
source
1. .787
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Athersys RAGE Library"
/cell_line="H1080"
/note="See 'Creation of Genome-wide Protein Expression
Libraries using Random Activation of Gene Expression',
method was used, these sequence tags are not necessarily
expressed in H1080 under normal circumstances."
BASE COUNT      230 a 177 c 142 g 226 t
ORIGIN

```



Nature Biotechnology, in press. Note that even though the cell type indicated is H1080, since a random activation method was used, these sequence tags are not necessarily expressed in H1080 under normal circumstances.

BASE COUNT 236 a 177 c 145 g 228 t 1 others

Query Match 26.8%; Score 250; DB 11; Length 787;  
Best Local Similarity 99.7%; Pred. No. 2.9e-67;  
Matches 300; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 165 tatcccggttactgacaaacccatccctccagatgagcgtccctaccctgcagaa 224  
|||||  
DB 77 TATCCGGTTACTGACAAACCCATCACCTTTCAGATGAGCGTCTTACCTGCAGAA 136  
|||||

QY 225 taactgtgtttatcaatttgggtctctccaccacacccctattctctggagagccctc 284  
|||||  
DB 137 TAACTGTGTTATCAATATTGGGCTCTTACCACACCCCTATTCTGGCAAGACCTTC 156  
|||||

QY 285 aggttttaaccccttgagattctccagggaataattctgaaaaatacatctatgctt 344  
|||||  
DB 197 AGGTCTTTAACCCCTTGAGATTCCTCAGGGAATAATCTGAAAAATATATCCCTATGCTT 256  
|||||

QY 345 tcataccattctcagctgagtaagaaactgcattgggcagcatttctggaataattgagt 404  
|||||  
DB 257 TCATACCATTCACGCTGGATTAAGAACTGCATTGGCAGCATTTTGCCATAATTGAGT 316  
|||||

QY 405 gtaagtgcagtgagcatttaactctcccttcgaagctgctccagaccactcaagcc 464  
|||||  
DB 317 GTAAGTGGCAGTGGCATTAATCTGCTCCGCTTCAAGCTGGCTCCACACACCTCAAGGC 376  
|||||

QY 465 c 465  
DB 377 c 377

RESULT 6  
H25624 432 bp mRNA EST 10-JUL-1995  
LOCUS y148g04.r1 Soares breast 3NBHst Homo sapiens cDNA clone  
DEFINITION IMAGE:161526 5' similar to gb:J02871 CYTOCHROME P450 1VBI (HUMAN  
); contains Alu repetitive element; mRNA sequence.

ACCESSION H25624  
VERSION H25624.1 GI:894747  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 432)  
AUTHORS Hillier, L., Clark, N., Dubouque, T., Elliston, K., Hawkins, M., Holman  
M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J.,  
Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevaskis, E., Waterston  
R., Williamson, A., Woldmann, P. and Wilson, R.

COMMENT The WashU-Merck EST Project  
Unpublished (1995)  
Contact: Wilton RK  
Washington University School of Medicine  
444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@wustl.wustl.edu

Insert Size: 874  
High quality sequence stops: 353  
Source: IMAGE Consortium, LLM  
This clone is available royalty-free through LLM; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Insert Length: 874 Std Error: 0.00  
Seq primer: M13Rpl  
High quality sequence stop: 353.

FEATURES  
source Location/Qualifiers  
1..432

/organism="Homo sapiens"  
/db\_xref="GDB:575683"  
/db\_xref="taxon:9606"  
/clone="IMAGE:161526"  
/clone\_lib="Soares breast 3NBHst"  
/sex="Female"  
/dev\_stage="adult"  
/lab\_host="DH10B (ampicillin resistant)"  
/notes="Organ: breast; Vector: pT73D (Pharmacia) with a  
modified polylinker; Site: 1; Site: 2; Eco RI; 1st  
strand cDNA was primed with a Not I - oligo(dT) primer [5'  
TGTAACCAATCTGAAGTGGAGCGCGCCCTTTTCTTTTCTTTT 3']  
double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not I  
and Eco RI sites of a modified pT73 vector (Pharmacia).  
Library went through one round of normalization to a Cot -  
20. Library constructed by Bento Soares and M.Fatima  
Bonaldi."

BASE COUNT 123 a 98 c 75 g 133 t 3 others

Query Match 26.5%; Score 247; DB 11; Length 432;  
Best Local Similarity 100.0%; Pred. No. 3.9e-66;  
Matches 247; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 219 caggaataactgtttatcaataattctgggtctctccaccacccctattcttgggaag 278  
|||||  
DB 85 CAGGAATAACTGTGTTTATCATATTTTGGCTCTTCCACCACACCCCTATTCTGGGAAG 144  
|||||

QY 279 accctcaggtctttaaaccctctgagattctccagggaataattctgaaaaatacatccct 338  
|||||  
DB 145 ACCCTCAGGTCTTTAACCCCTTGAGATTCCTCAGGGAATAATCTGAAAAATATCATCCCT 204  
|||||

QY 339 atgcttctaccattctcagctggattgaaggaactgcatl3ggcagcattttgcataa 398  
|||||  
DB 205 ATGCTCTTACCATTCACGCTGGATTAAGGAATTCATTCGCGCATTTTGCCATAA 264  
|||||

QY 399 ttgagtgaagtggcagtggtggaattactctgctcagcttcaagctcagaccact 458  
|||||  
DB 265 TTGAGTGAAGTGGCAGTGGCATTAACCTGCTCCCTTCAAGCTGGCTCCAGACCACCT 324  
|||||

QY 459 caagggc 465  
DB 325 CAAGGCC 331  
|||||

RESULT 7  
A1733538 535 bp mRNA EST 24-OCT-2000  
LOCUS y148g04.y5 Soares breast 3NBHst Homo sapiens cDNA clone  
DEFINITION IMAGE:161526 5' similar to SW:CP4B\_RABII P15128 CYTOCHROME P450 4B1  
; mRNA sequence.

ACCESSION A1733538  
VERSION A1733538.1 GI:5054699  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 535)  
AUTHORS NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT Other ESTs: y148g04.x5  
Contact: Robert Strausberg, Ph.D.  
Email: cgafts@mail.nih.gov  
This clone is available royalty-free through LLM; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
This read is a RESEQUENCE of a previously sequenced human clone  
Original clone citation: see original entry for original citation  
information

This read has been verified (found to hit its original self in the correct orientation)  
Insert Length: 874 Std Error: 0.00  
Seq primer: -4ORP from Gibco  
High quality sequence stop: 440.

FEATURES  
source

1. 535  
/organism="Homo sapiens"  
/db\_xref="GDB:575683"  
/db\_xref="taxon:9606"  
/clone="IMAGE:161526"  
/clone\_lib="Soares breast 3NBHst"  
/sex="Female"  
/dev\_stage="adult"  
/lab\_host="DH10B (ampicillin resistant)"  
/note="Organ: breast; Vector: pT7T3D (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGAGCGGCCCTTTTITTTTTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT7T3 vector (Pharmacia). Library went through one round of normalization to a Cot = 20. Library constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 175 a 109 c 90 g 161 t  
ORIGIN

Query Match 26.5%; Score 247; DB 10; Length 535;

Best Local Similarity 100.0%; Pred. No. 3.3e-66; Mismatches 0; Indels 0; Gaps 0;

Matches 247; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 219 caggataaactgtttatcaatatttgggtctctccacacacccctattttctgggaag 278  
DB 85 CAGGAATAACTGTGTTATCAATATTGGGCTCTCCACCAACCCCTATTTCGGGAAG 144  
QY 279 accctcaggtcttaaccccttgagattctccagggaataattctgaaaaatacatccct 338  
DB 145 ACCCTCAGGTCTTTAACCCCTTGAGATCTCCAGGGAAAAATTCGAAAAATACATCCCT 204  
QY 339 atgcttcataccattctcagctggattgaaggaactgcatggcgagcattttgccataa 396  
DB 205 ATGCCCTCATACCATCTCAGCTGGATTAGGAACATGCAATGGCAGCATTTTGCCTAAT 264  
QY 399 ttgagtgtaagtgccagtggaactaactctgtccgcttcagagctggctccagaccact 458  
DB 265 TTGAGTGTAAGTGCGAGTGGCATTAACCTCTGCTCCGCTTCAAGCTGGCTCCAGACCAC 324  
QY 459 caagccc 465  
DB 325 CAAGGCC 331

RESULT 8  
AT820775 518 bp mRNA EST 21-OCT-2000  
LOCUS AT820775.1 GI:5439854  
DEFINITION Y138c11.y5 Soares breast 3NBHst Homo sapiens cDNA clone IMAGE:160332.5, similar to SW:CP4B\_RABIT P15128 CYTOCHROME P450 451 ;, mRNA sequence.

ACCESSION AT820775  
VERSION AT820775.1  
KEYWORDS EST.

SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1 (bases 1 to 548)  
AUTHORS NCI-CCGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT Other\_ESTs: y138c11.x5

Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
This clone is available royalty-free through, LNL : contact the IMAGE Consortium (info:image.lnl.gov) for further information.  
This read is a RESEQUENCE of a previously sequenced human clone  
Original clone citation: see original entry for original citation  
Information  
This read has been verified (found to hit its original self in the correct orientation)  
Insert Length: 870 Std Error: 0.00  
Seq primer: -4ORP from Gibco  
High quality sequence stop: 469.

FEATURES  
source

1. 518  
/organism="Homo sapiens"  
/db\_xref="GDB:574579"  
/db\_xref="taxon:9606"  
/clone="IMAGE:160532"  
/clone\_lib="Soares breast 3NBHst"  
/sex="Female"  
/dev\_stage="adult"  
/lab\_host="DH10B (ampicillin resistant)"  
/note="Organ: breast; Vector: pT7T3D (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGAGCGGCCCTTTTITTTTTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT7T3 vector (Pharmacia). Library went through one round of normalization to a Cot = 20. Library constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 180 a 111 c 94 g 163 t  
ORIGIN

Query Match 26.5%; Score 247; DB 10; Length 548;

Best Local Similarity 100.0%; Pred. No. 3.2e-66; Mismatches 0; Indels 0; Gaps 0;

Matches 247; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 219 caggataaactgtttatcaatatttgggtctctccacacacccctattttctgggaag 278  
DB 99 CAGGAATAACTGTGTTATCAATATTGGGCTCTCCACCAACCCCTATTTCGGGAAG 158  
QY 279 accctcaggtcttaaccccttgagattctccagggaataattctgaaaaatacatccct 338  
DB 159 ACCCTCAGGTCTTTAACCCCTTGAGATCTCCAGGGAAAAATTCGAAAAATACATCCCT 218  
QY 339 atgcttcataccattctcagctggattgaaggaactgcatggcgagcattttgccataa 398  
DB 219 ATGCCCTCATACCATCTCAGCTGGATTAGGAACATGCAATGGCAGCATTTTGCCTAAT 278  
QY 399 ttgagtgtaagtgccagtggaactaactctgtccgcttcagagctggctccagaccact 458  
DB 275 TTGAGTGTAAGTGCGAGTGGCATTAACCTCTGCTCCGCTTCAAGCTGGCTCCAGACCAC 338  
QY 459 caagccc 465  
DB 339 CAAGGCC 345

RESULT 9  
BG192933

LOCUS BG192933 767 bp mRNA EST 21-APR-2001  
DEFINITION RST12058 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.  
ACCESSION BG192933  
VERSION BG192933.1 GI:13714620  
KEYWORDS EST.  
SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1 (bases 1 to 767)

JOURNAL.  
COMMENT

**AUTHORS**

Harrington, J. J., Sherf, B., Rundlett, S., Jackson, P. D., Perry, R., Cain, S., Dahl, T., Thornton, M., Kamachandran, R., Whittington, J., Lerner, L., Krashoc, D., McElligott, K., Clark, S., Ways, R., Smith, E., Veloso, N., Jless, J., Cothren, K., Lo, K., Offenbacher, J., Danzig, J., and Ducar, M. *Gene* expression libraries using Random

**AUTHORS**  
**TITLE**  
**JOURNAL**  
**COMMENT**

NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP).  
 Tumor Gene Index  
 Unpublished (1997)  
 Contact: Robert Strausberg, Ph.D.  
 Email: [cgaps@mail.nih.gov](mailto:cgaps@mail.nih.gov)  
 This clone is available royalty-free through LINL; contact the  
 IMAGE Consortium ([infoimage.lnl.gov](http://infoimage.lnl.gov)) for further information.  
 This read is a RESEQUENCE of a previously sequenced human clone  
 Original clone citation: see original entry for original citation  
 information  
 This read has been verified (found to hit its original self in the  
 correct orientation)  
 Possible reversed clone: similarity on wrong strand  
 Insert length: 874 Std Error: 0.00  
 Seq primer: -400P from Gibco  
 High quality sequence stop: 460.

**FEATURES**  
 source

1..629  
 /organism="Homo sapiens"  
 /db\_xref="GDB:575683"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:161526"  
 /clone\_lib="Soares breast 3NDBst"  
 /sex="Female"  
 /dev\_stage="adult"  
 /lab\_host="DH10B (ampicillin resistant)"  
 /note="Organ: breast; Vector: p773D (Pharmacia) with a  
 modified polylinker; Site\_1: Not I; Site\_2: Eco RI; lstr  
 strand cDNA was primed with a Not I - oligo(dT) primer [5',  
 TGTTACCAATCTGAAGTGGAGCGCCGCTTTTCTTTTCTTTT 3'],  
 double-stranded cDNA was ligated to Eco RI adaptors  
 (Pharmacia), digested with Not I and cloned into the Not I  
 and Eco RI sites of a modified p773 vector (Pharmacia).  
 Library went through one round of normalization to a Cot =  
 20. Library constructed by Bento Soares and M.Fatima  
 Bonaldo."

BASE COUNT 196 a 106 c 127 g 196 t 4 others

**Query Match** 20.0%; Score 187; DB 10; Length 629;  
**Best Local Similarity** 100.0%; Pred. No. 6.1e-48;  
**Matches** 187; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 564 gagacaattttctaccagaagaacacaaaggataataatacaaaaatatatgat 623  
 |||||||  
 LL 358 GAGACAAATTTCTCTACCAAGGAGAGACAAAGGATAATAATACAAATATATGAT 299  
 |||||||  
 QY 624 atggtttgtgacaaattataaacttaggatactctgactggttttgacatccattaa 683  
 |||||||  
 LL 298 ATGGTGTGTGACAAATATATAACTTAGGATCTTCGACTGGTTTGACATCCATTA 239  
 |||||||  
 QY 684 cagtaatttttaattttctgctgtatctggtgaaacccacacaaacacactgaaaaaactc 743  
 |||||||  
 DL 238 CAGTAATTTTAATTTCTTCTGTCTATCTGTTGAAACCCACCAAAACACCTGAAAAAATC 179  
 |||||||

QY 744 aagctga-750  
 |||||||  
 DL 178 AAGCTGA-172

**RESULT** 12  
**LOCUS** AV700083 760 bp mRNA EST 25-SEP-2000  
**DEFINITION** AV700083 GK Homo sapiens CDNA clone GKCBF04 3', mRNA sequence.  
**ACCESSION** AV700083  
**VERSION** AV700083.1 GI:10302054  
**KEYWORDS** EST.  
**SOURCE** human.  
**ORGANISM** Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.

**REFERENCE**  
**AUTHORS**  
**TITLE**  
**JOURNAL**  
**COMMENT**

1 (bases 1 to 760)  
 Wu, F., Qian, B., Huang, C., Huang, C., Kang, B., Gao, X., Xu, Z., Xiao, H.,  
 Xu, X., Li, N., Peng, Y., Liu, F., Qu, J., Song, H., Cheng, Z., Zeng, L.,  
 Xu, S., Gu, W., Tu, Y., Jia, J., Fu, G., Ren, S., Zhong, M., Lu, G., Yang,  
 Y., Gu, Y., Chen, Z., and Han, Z.  
 Homo sapiens CDNA GK- clones  
 Unpublished (2000)  
 Contact: Zengqiang Han  
 Chinese National Human Genome Center at Shanghai  
 331 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai  
 201203, P. R. China  
 Tel: 86-21-50801519 (ex.45)  
 Fax: 86-21-50801922  
 Email: [hanzq@chgc.sh.cn](mailto:hanzq@chgc.sh.cn)  
 This clone is available at CHGC in Shanghai.

**FEATURES**  
 Location/Qualifiers

1..760  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="GKCBF04"  
 /clone\_lib="GK"  
 /tissue\_type="hepatocellular carcinoma"  
 /dev\_stage="adult"  
 /lab\_host="SOLR"  
 /note="Vector: pBluescript sk(-); Site\_1: EcoRI; Site\_2:  
 XhoI"

BASE COUNT 235 a 136 c 164 g 224 t 1 others

**Query Match** 20.0%; Score 187; DB 10; Length 760;  
**Best Local Similarity** 100.0%; Pred. No. 5.2e-48;  
**Matches** 187; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 564 gagacaattttctaccagaagaacacaaaggataataatacaaaaatatatgat 623  
 |||||||  
 DL 334 GAGACAAATTTCTCTACCAAGGAGAGACAAAGGATAATAATACAAATATATGAT 275  
 |||||||  
 QY 624 atggtttgtgacaaattataaacttaggatactctgactggttttgacatccattaa 683  
 |||||||  
 DL 274 ATGGTGTGTGACAAATATATAACTTAGGATCTTCGACTGGTTTGACATCCATTA 215  
 |||||||  
 QY 684 cagtaatttttaattttctgctgtatctggtgaaacccacacaaacacactgaaaaaactc 743  
 |||||||  
 DL 214 CAGTAATTTTAATTTCTTCTGTCTATCTGTTGAAACCCACCAAAACACCTGAAAAAATC 155  
 |||||||

QY 744 aagctga-750  
 |||||||  
 DL 154 AAGCTGA-148

**RESULT** 13  
**LOCUS** BG196956/c 858 bp mRNA EST 21-APR-2001  
**DEFINITION** RSI16188 Athysys RAGE Library Homo sapiens CDNA, mRNA sequence.  
**ACCESSION** BG196956  
**VERSION** BG196956.1 GI:13718643  
**KEYWORDS** EST.  
**SOURCE** human.  
**ORGANISM** Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.

**REFERENCE**  
**AUTHORS**  
**TITLE**  
**JOURNAL**  
**COMMENT**

1 (bases 1 to 858)  
 Harrington, J. J., Sherf, B., Rundlett, S., Jackson, P. D., Perry, R.,  
 Cain, S., Dahl, T., Thornton, M., Ramachandran, K., Whittington, J.,  
 Ierner, L., Krashoc, D., McElligott, K., Clark, S., Mays, R., Smith, E.,  
 Veloso, N., Hess, J., Cothren, K., Lo, K., Offenbacher, J., Danzig, J.,  
 and Ducar, M.  
 Creation of Genome-wide Protein Expression Libraries using Random  
 Activation of Gene Expression  
 Nat. Biotechnol. 19 (5), 440 (2001) In press  
 Contact: Scott J. Cain  
 Athysys, Inc





```

RESULT 18
BG212152 807 bp mRNA EST 21-APR-2001
LOCUS
DEFINITION RST31623 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.
ACCESSION BG212152
VERSION BG212152.1 GI:13733727
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 807)
Harrington, J.J., Sherf, B., Rundlett, S., Jackson, P.D., Perry, R.,
Cain, S., Dahl, T., Thornton, M., Ramachandran, R., Whittington, J.,
Lerner, L., Krashinsky, D., McElligott, K., Clark, S., Mays, R., Smith, E.,
Veloso, N., Hess, J., Colhoun, K., Lo, K., Offenbacher, J., Danzig, J.,
and Nucifora, M.
Creation of Genome-wide Protein Expression Libraries using Random
Activation of Gene Expression
Nat. Biotechnol. 19 (5), 440 (2001) In press
Contact: Scott J. Cain
Athersys, Inc.
3201 Carnegie Ave., Cleveland, OH 44115, USA
Tel: 216 431 9900
Fax: 216 361 9596
Email: scain@atersys.com
High quality sequence stop: 465.
Location/Qualifiers
1..807
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Athersys RAGE Library"
/cell_line="HT1080"
/Note="See 'Creation of Genome-wide Protein Expression
Libraries using Random Activation of Gene Expression',
Nature Biotechnology, in press. Note that even though the
cell type indicated is HT1080, since a random activation
method was used, these sequence tags are not necessarily
expressed in HT1080 under normal circumstances."
BASE COUNT 230 a 200 c 164 g 212 t 1 others
ORIGIN
Query Match 18.5%; Score 173; DB 11; Length 807;
Best Local Similarity 99.6%; Pred. No. 9,3e-44;
Matches 223; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 242 attgggctctccaccacacccctattctgggaagaccctcagylcttttaaccccttg 301
|||||
Db 284 ATTGGGCTCTTCCACCACACCCCTATTCTGGGAAGACCCTCAGGCTTTAAACCCCTTG 343
|||||
QY 302 agattctccagggaattctgaaaaatacatccctatgccttcataccattctcagct 361
|||||
Db 344 AGATTCTCCAGGGGAAATTCGAAAATAACATCCCTATGCCTTCATACCACTACGCT 403
|||||
QY 362 ggattaaggaaactgcattggcgsgcattttggccataatgagtgtaaaagtgccagcgca 421
|||||
Db 404 GGATTAAGGAAGTGCATTGGGAGACATTTTGGCCATAATGAGTGAAGTGGCAGTGCCA 463
|||||
QY 422 ttaactctgctccgcttcaagctggtccagaccactcaaggcc 465
|||||
Db 464 TTAACCTGCTCCGCTTCAAGCTGGCTCCAGACCACCTCAAGGCC 507
|||||

RESULT 19
AW472756/c 239 bp mRNA EST 24-FEB-2000
LOCUS
DEFINITION x019a03.x1 NCI_CGAP_U11 Homo sapiens cDNA clone IMAGE#2751052 3',
mRNA sequence.
ACCESSION AW472756
VERSION AW472756.1 GI:7042862

```

```

KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 239)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael K.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone Distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Seq primer: -40UP from Gibco
High quality sequence stop: 223.
Location/Qualifiers
1..239
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2751052"
/clone_lib="NCI_CGAP_U11"
/tissue_type="well-differentiated endometrial
adenocarcinoma, 7 pooled tumors"
/lab_host="DH10B"
/Note="Organ: uterus; Vector: pCMV-SPOK6; Site: 1; Sali;
Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.75 kb. Life technologies catalog #
11538-014"
BASE COUNT 79 a 36 c 54 g 70 t
ORIGIN
Query Match 14.8%; Score 138; DB 10; Length 239;
Best Local Similarity 100.0%; Pred. No. 1.2e-32;
Matches 138; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 782 gtaactagtgtagtggtttcaagcagttttgatcaaaacctcactcagtatctgc 841
|||||
Db 139 GTAACAGTGTAGAGTGCGCTTTCAGCAAGATTGTGATCAAAACCCACACAGTATCTGC 80
|||||
QY 842 attacttttatctctgcaaatatctgcatgagcttctatctcagttatcttcccat 901
|||||
Db 79 ATTACTTTTATCTCTGCAAAATATCTGCATGATGATGATTTATCTTATCTTCTCCCAT 20
|||||
QY 502 aataaaaaatatctgcca 919
|||||
Db 15 AATAAAAAATATCTGCCA 2
|||||

RESULT 20
H21976 332 bp mRNA EST 06-JUL-1995
LOCUS
DEFINITION Y136c11.r1 Soares breast 3nbH8st Homo sapiens cDNA clone
IMAGE:160532 5', similar to SP:CP4B_RABIT P15128 CYTOCHROME P450
IVB1 ;, mRNA sequence.
ACCESSION H21976
VERSION H21976.1 GI:890671
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 332)
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman
M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J.,
Rifkin, L., Rohlffing, T., Soares, M., Tan, F., Trevaskis, E., Waterston

```

Tue Apr 30 14:18:38 2002

R., Williamson, A., Wohlmann, P. and Wilson, R.  
The WashU-Merck EST Project  
Unpublished (1995)  
Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
Insert Size: 870  
High quality sequence steps: 233  
Source: IMAGE Consortium, LLNL  
This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Insert Length: 870 Std Error: 0.00  
Seq primer: M13Rpl  
High quality sequence stop: 233.

## FEATURES

Source  
1. .332  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="GDB:574579"  
/db\_xref="taxon:9606"  
/clone="IMAGE:160532"  
/clone\_lib="Soares breast 3NBHbst"  
/sex="Female"  
/dev\_stage="adult"  
/lab\_host="DH10B (ampicillin resistant)"  
/note="Organ: breast; Vector: pT73D (Pharmacia) with a  
modified polylinker; Site\_1: Not 1; Site\_2: Eco RI; 1st  
strand cDNA was primed with a Not I - oligo(dT) primer [5',  
TCTTACCAATCTGAAGTGGAGCGCGCCCTTTTCTTTTCTTTT 3'],  
double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not I  
and Eco RI sites of a modified pT73 vector (Pharmacia).  
Library went through one round of normalization to a Cot =  
20. Library constructed by Bento Soares and M.Fatima  
Bonaldo."

BASE COUNT 80 a 88 c 58 g 102 t 4 others  
ORIGIN

Query Match 14.6%; Score 136; DB 11; Length 332;  
Best Local Similarity 100.0%; Pred. No. 3.9e-32;  
Matches 136; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 219 caggataactgtttatcaatttgggtctctccaccacacccctatttctgggaag 278  
|||||  
Db 44 CAGGAATACTGTGTTTATCAATATTTGGCTCTTCCACCACACCCCTATTTCTGGGAAG 103  
QY 279 accctcagggttttaaccccttgagattctccagggaataattctgaaaaatacatcctc 338  
|||||  
Db 104 ACCCTCAGGTCTTTAAACCCCTTGAGATCTCCAGGGAAATTCGAAAAAATACATCCCT 163  
QY 339 atgccttcataccatt 354  
Db 164 ATGCCCTTCATACCAT 179

## RESULT 21

H25577/c 421 bp mRNA EST 10-JUL-1995  
LOCUS Y148g04.s1 Soares breast 3NBHbst Homo sapiens cDNA clone  
DEFINITION IMAGE:161526 3', mRNA sequence.  
ACCESSION H25577  
VERSION H25577.1 GI:894700  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 421)  
AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J.,

TITLE  
JOURNAL  
COMMENT

Rifkin, L., Rohlfing, T., Soares, M., Ian, F., Irevaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.  
The WashU-Merck EST Project  
Unpublished (1995)  
Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
Insert Size: 874  
High quality sequence steps: 297  
Source: IMAGE Consortium, LLNL  
This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Insert Length: 874 Std Error: 0.00  
Seq primer: Promega -21ml3  
High quality sequence stop: 297.

## FEATURES

Source  
1. .421  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="GDB:575683"  
/db\_xref="taxon:9606"  
/clone="IMAGE:161526"  
/clone\_lib="Soares breast 3NBHbst"  
/sex="Female"  
/dev\_stage="adult"  
/lab\_host="DH10B (ampicillin resistant)"  
/note="Organ: breast; Vector: pT73D (Pharmacia) with a  
modified polylinker; Site\_1: Not 1; Site\_2: Eco RI; 1st  
strand cDNA was primed with a Not I - oligo(dT) primer [5',  
TCTTACCAATCTGAAGTGGAGCGCGCCCTTTTCTTTTCTTTT 3'],  
double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not I  
and Eco RI sites of a modified pT73 vector (Pharmacia).  
Library went through one round of normalization to a Cot =  
20. Library constructed by Bento Soares and M.Fatima  
Bonaldo."

BASE COUNT 132 a 66 c 80 g 139 t 4 others  
ORIGIN

Query Match 13.3%; Score 124; DB 11; Length 421;  
Best Local Similarity 100.0%; Pred. No. 1.5e-28;  
Matches 124; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 782 gtaactagtgtagagtggtttcaagcatagtttgatcaaaactccactcagtatctgc 841  
|||||  
Db 125 GTAAC TAGTG TAGTG GCGCTTTCACACATAGTTTGATCAAAACTCCACTCAGTATCTGC 66  
QY 842 attcttttattctctgcaaatattctgcattgatagctttattctcagttattcttccccat 901  
|||||  
Db 65 ATTACTTTTACTCTGCAAAATATCTGCATGATAGTAGCTTTATTCAGTATCTCTTTCCTCCAT 6  
QY 902 aata 905  
Db 5 AATA 2

## RESULT 22

H21577/c 444 bp mRNA EST 06-JUL-1995  
LOCUS Y138c11.s1 Soares breast 3NBHbst Homo sapiens cDNA clone  
DEFINITION IMAGE:160532 3', mRNA sequence.  
ACCESSION H21577  
VERSION H21577.1 GI:890672  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 444)  
AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman





```
/db_xref="taxon:9606"
/clone_lib="Athersys RAGE Library"
/cell_line="HT1080"
/notes="See 'Creation of Genome-wide Protein Expression
Libraries using Random Activation of Gene Expression',
Nature Biotechnology, in press. Note that even though the
cell type indicated is HT1080, since a random activation
method was used, these sequence tags are not necessarily
expressed in HT1080 under normal circumstances."

BASE COUNT 69 a 68 c 58 g 57 t
ORIGIN

Query Match 12.6%; Score 118; DB 11; Length 252;
Best Local Similarity 100.0%; Pred. No. 1.5e-26;
Matches 118; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ttactgttgcgaagtaccctgagcatcagcagagatgccgagatgaaatcagggaac 60
|||||
Db 107 TTTACTGCTTGGCAAGTACCTGACATCAGCAGATGCCGAGATGAAATCAGGGAC 166
|||||

Qy 61 tctaggggatgggtcttattaccctgggaacacctgagcagatgccttacacac 118
|||||
Db 167 TCCTAGGGATGGGTCTTCTATTACCTGGGAACCTGAGCCAGATGCCTTACACCAC 224
|||||

RESULT 25
BG219912/c 240 bp mRNA EST 21-APR-2001
LOCUS
DEFINITION RST39683 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.
ACCESSION BG219912
VERSION BG219912.1 GI:13745933
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 240)
AUTHORS Harrington,J.J., Sherf,B., Rundlett,S., Jackson,P.D., Perry,R.,
Cain,S., Dahl,T., Thornton,M., Ramachandran,R., Whittington,J.,
Lerner,L., Krashoc,D., McElligott,K., Clark,S., Mays,R., Smith,E.,
Veloso,N., Hess,J., Cotiren,K., Lo,K., Offenbacher,J., Danzig,J.
and Ducar,M.
TITLE Creation of Genome-wide Protein Expression Libraries using Random
Activation of Gene Expression
JOURNAL Nat. Biotechnol. 19 (5), 440 (2001) In press
COMMENT Contact: Scott J. Cain
Athersys, Inc.
3201 Carnegie Ave, Cleveland, OH 44115, USA
Tel: 216 431 9900
Fax: 216 361 9596
Email: scain@athersys.com
High quality sequence stop: 190.
FEATURES
Location/Qualifiers
source 1..240
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Athersys RAGE Library"
/cell_line="HT1080"
/notes="See 'Creation of Genome-wide Protein Expression
Libraries using Random Activation of Gene Expression',
Nature Biotechnology, in press. Note that even though the
cell type indicated is HT1080, since a random activation
method was used, these sequence tags are not necessarily
expressed in HT1080 under normal circumstances."

BASE COUNT 85 a 40 c 53 g 62 t
ORIGIN

Query Match 11.6%; Score 108; DB 11; Length 240;
Best Local Similarity 100.0%; Pred. No. 1.8e-23;
Matches 108; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 85 a 40 c 53 g 62 t
|||||
Db 147 ACCGACCCGCTAGTAAAC 163
|||||

Qy 148 acgcacggtagtaaac 164
|||||
Db 147 ACCGACCCGCTAGTAAAC 163
|||||

Query Match 8.3%; Score 77; DB 11; Length 267;
Best Local Similarity 100.0%; Pred. No. 4.8e-14;
Matches 77; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 86 ggaacacctgagctagatgccttacaccacgatgcatcagaagatgccctccctct 147
|||||
Db 87 GGGACACCTGAGCCAGATGCCCTTACACCACCATGTCATCAAGGAATGCCCTCCCTCT 146
|||||

Qy 148 acgcacggtagtaaac 164
|||||
Db 147 ACCGACCCGCTAGTAAAC 163
|||||

RESULT 27
AQ505963 729 bp DNA GSS 29-APR-1999
LOCUS
DEFINITION RPCI-11-316A11.TJ RPCI-11 Homo sapiens genomic clone RPCI-11-316A11
DNA sequence.
ACCESSION AQ505963
VERSION AQ505963.1 GI:4710710
KEYWORDS GSS.
```

SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

Human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 729)  
Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,K. and Venter  
J.C.  
Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready  
Map Building  
Contact: Shaying Zhao, William Nierman, Mark Adams  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: hbeetigr.org  
Clones are derived from the human BAC library RPCI-11. For BAC  
library availability, please contact Pieter de Jong  
(pieter@jeong.med.buffalo.edu). Clones may be purchased from  
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering\_bac.htm)  
or from Research Genetics (info@resgen.com). BAC end search page:  
http://www.tigr.org/tdb/humgen/bac\_end\_search/bac\_end\_search.html.  
Seq primer: SP6  
Class: BAC ends.

FEATURES  
source  
Location/Qualifiers  
1..729  
/organism="Homo sapiens"  
/db\_xref="GDB:7620970"  
/db\_xref="taxon:9606"  
/clone="RPCI-11-316A11"  
/clone\_lib="RPCI-11"  
/sex="Male"  
/cell\_type="Lymphocytes"  
/note="Vector: pBAC3.6; Site\_1: EcoRI; Site\_2: EcoRI;  
RPC111 Human Male BAC Library"  
BASE COUNT 200 a 166 c 137 g 226 t  
ORIGIN

Query Match 8.3%; Score 77; DB 13; Length 729;  
Best Local Similarity 100.0%; Pred. No. 2.1e-14;  
Matches 77; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 88 gggaacacctgagcagatgcttacaccacgagtgatcacaaggaatgctcgcctct 147  
|||||  
Db 499 GGGACACCTGAGCCAGATGCTTACACCACGATGTCATCAAGGATGCTCGCTCT 559  
|||||

QY 148 acgacccgtagtaaac 164  
|||||  
Db 559 ACCACCGTAGTAAC 575  
|||||

RESULT 28  
A0450557  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
COMMENT

A0450557 498 bp DNA GSS 21-APR-1999  
HS\_5166.AL.E01.T7A.RPCI-11 Human Male BAC Library  
genomic clone Plate-742 Col-1 Row-1, DNA sequence.  
A0450557  
GSS.  
A0450557.1 GI:4592013  
human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 498)  
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,  
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and  
Hood,L.  
Sequence-tagged connectors: A sequence approach to mapping and  
scanning the human genome  
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)  
99380589

SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
COMMENT

Human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 742)  
Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,K. and Venter  
J.C.  
Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready  
Map Building  
Contact: Shaying Zhao, William Nierman, Mark Adams  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: hbeetigr.org  
Clones are derived from the human BAC library RPCI-11. For BAC  
library availability, please contact Pieter de Jong  
(pieter@jeong.med.buffalo.edu). Clones may be purchased from  
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering\_bac.htm)  
or from Research Genetics (info@resgen.com). BAC end search page:  
http://www.tigr.org/tdb/humgen/bac\_end\_search/bac\_end\_search.html.  
Seq primer: SP6  
Class: BAC ends.

FEATURES  
source  
Location/Qualifiers  
1..498  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="Plate-742 Col-1 Row-1"  
/clone\_lib="RPCI-11 Human Male BAC Library"  
/sex="male"  
/note="Vector: pBAC3.6; Site\_1: EcoRI; Site\_2: EcoRI;  
Male blood DNA was isolated from one randomly chosen donor  
and partially digested with a combination of EcoRI and  
EcoRI Methylase. Size selected DNA was cloned into the  
pBAC3.6 vector at EcoRI sites"  
BASE COUNT 118 a 109 c 112 g 155 t 4 others  
ORIGIN

Query Match 6.2%; Score 58; DB 13; Length 498;  
Best Local Similarity 100.0%; Pred. No. 1.8e-06;  
Matches 58; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 313 ggaaattctgaaatacatcctctatgcttcattcattcattcagctgattaaag 370  
|||||  
Db 149 GGAATATCTGAAAAATACATCCCTATGCTTACCATCTCAGCTGGATTAAAG 206  
|||||

RESULT 25  
A0430949/c  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
COMMENT

A0430949 521 bp DNA GSS 31-MAR-1999  
HS\_5137.AL.H03.T7A.RPCI-11 Human Male BAC Library  
genomic clone Plate-713 Col-5 Row-O, DNA sequence.  
A0430949  
GSS.  
A0430949.1 GI:4541284  
human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 521)  
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,  
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and  
Hood,L.  
Sequence-tagged connectors: A sequence approach to mapping and  
scanning the human genome  
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)  
99380589

SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
COMMENT

Human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 742)  
Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,K. and Venter  
J.C.  
Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready  
Map Building  
Contact: Shaying Zhao, William Nierman, Mark Adams  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: hbeetigr.org  
Clones are derived from the human BAC library RPCI-11. For BAC  
library availability, please contact Pieter de Jong  
(pieter@jeong.med.buffalo.edu). Clones may be purchased from  
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering\_bac.htm)  
or from Research Genetics (info@resgen.com). BAC end search page:  
http://www.tigr.org/tdb/humgen/bac\_end\_search/bac\_end\_search.html.  
Seq primer: SP6  
Class: BAC ends.

FEATURES  
source  
Location/Qualifiers  
1..498  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="Plate-742 Col-1 Row-1"  
/clone\_lib="RPCI-11 Human Male BAC Library"  
/sex="male"  
/note="Vector: pBAC3.6; Site\_1: EcoRI; Site\_2: EcoRI;  
Male blood DNA was isolated from one randomly chosen donor  
and partially digested with a combination of EcoRI and  
EcoRI Methylase. Size selected DNA was cloned into the  
pBAC3.6 vector at EcoRI sites"  
BASE COUNT 118 a 109 c 112 g 155 t 4 others  
ORIGIN

Query Match 6.2%; Score 58; DB 13; Length 498;  
Best Local Similarity 100.0%; Pred. No. 1.8e-06;  
Matches 58; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 313 ggaaattctgaaatacatcctctatgcttcattcattcattcagctgattaaag 370  
|||||  
Db 149 GGAATATCTGAAAAATACATCCCTATGCTTACCATCTCAGCTGGATTAAAG 206  
|||||

RESULT 25  
A0430949/c  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
COMMENT

A0430949 521 bp DNA GSS 31-MAR-1999  
HS\_5137.AL.H03.T7A.RPCI-11 Human Male BAC Library  
genomic clone Plate-713 Col-5 Row-O, DNA sequence.  
A0430949  
GSS.  
A0430949.1 GI:4541284  
human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 521)  
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,  
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and  
Hood,L.  
Sequence-tagged connectors: A sequence approach to mapping and  
scanning the human genome  
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)  
99380589

us-09-248-178-55.rst

/clone\_lib="UI-R-CAL"  
 /lab\_host="DL06 (Life Technologies)"  
 /note="Vector: pT7D3-pac (Pharmacia) with a modified  
 polylinker: Site\_1: Not I; Site\_2: Eco RI; The UI-R-CAL  
 library is a subtracted library derived from the following  
 tissues: thalamus, cerebellum, hypothalamus, medulla, pons  
 , midbrain, cerebral cortex, corpus striatum, testis, and  
 hippocampus. For a detailed description of the library  
 from which this clone was derived, please visit our web  
 site at [rateat.eng.utoro.edu](http://rateat.eng.utoro.edu). The subtraction has been  
 previously described in (Bonaldo, Lennon and Soares,  
 Genome Research 6:791-806, 1996)  
 TAG\_LIB=UI-R-CAL  
 TAG\_TISSUE=hippocampus  
 TAG\_SEQ=CATTG

midbrain, cerebral cortex, corpus striatum, the library hippocampus. For a detailed description of the library from which this clone was derived, please visit our web site at [atlas.csi.cmu.edu](http://atlas.csi.cmu.edu). The subtraction has been previously described in (Bonaldo, Lennon and Soares, 1998) (Accession # Genbank: U6791-806, 1996).  
TAG: 18B-UT-R-CAL  
TAG: TISSUE=hippocampus  
TAG: SEQ=CATIG.\*

TAG SEO=GATIG

Query Match	2.9%	Score 27;	DB 11;	Length 347;
Best Local Similarity	100.0%	Pred. No. 73;		
Matches	27;	Conservative 0;	Mismatches 0;	Indels 0;
Gaps	0;			

Oy	907 aaataatctgcccaaaaaaaaaaaaaaa 933   25 aaatatctgccaAAAAAAAAAAAAAA 3
D6	

RESULT 31  
55043410

RESULT	31	EST	10-OCT-2000
BF043410			
LOCUS			
DEFINITION			
	BF043410	483 bp	mRNA
	Bp250006B20D12	Soares	normalized bovine placenta
	clone Bp250006B20D12	5'	mRNA sequence.

ACCESSION BF043410  
VERSION BF043410.1 GI:10760465  
KEYWORDS EST.

SOURCE	ORGANISM	COW.
	<i>Bos taurus</i>	
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;	
	Bovidae; Bovinae; Bos.	

REFERENCE	1 (bases 1 to 483)
AUTHORS	Lewin, H.A., Soares, M.B., Rebeiz, M., Pardini, J., Liu, L. and Larson, J.H.
TITLE	MOVING ESTS

TITLE	JOURNAL	COMMENT
Bovine ESTs	Unpublished (2000)	
Contact: Lewin, H. A.		

W. M. Reek Center for Comparative and Functional Genomics  
University of Illinois at Urbana-Champaign  
340 Edward R. Madigan Laboratory, 1201 W. Gregory Dr., Urbana, IL  
61801, USA

Tel: 217 333 5998  
Fax: 217 244 5617  
Email: h-lewin@uiuc.edu  
Sequencing for cattle FAST sequencing was provided by the USDA National

Research Initiative, Animal Genome Resource Grant AG 99-95334 to H. A. Lewin and J. E. Womack. Base Calling/Quality Scores: PHREASE from Washington University Genome Center. Vector Trimmi g: from Washington University Genome Center. pubseq suite: from Washington University Genome Center. pubseq suite:

Cross-match from Washington University Genome Center prior to subcloning. This sequence is vector free and at least 200 bp in length.

PCR Primers

FORWARD: TAATACGACTCACTATACGG

BACKWARD: ATTAAACCTCACTAAG  
Insert Length: 483 Std Error: 0.00  
Plate: EP250006B20 row: D column: 12  
seq primer: AGCGGATACAAATTTCACACAGGA

FEATURES	Location/Qualifiers
source	1. .483

```

/organism="Bos taurus"
/db_xref="taxon:9913"
/clone="BP20006B20D12"
/clone.lib="Soares normalized bovine placenta"

```

us-09-248-178-55.rst

Tue Apr 30 14:18:38 2002

ORIGIN

```

Query Match      2.8%; Score 26; DB 11; Length 483;
Best Local Similarity 100.0%; Pred. No. 1:1e+02;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 200 gatggagcgtcttaccctgcaggaaat 225
      |||||||
DB 298 GATGGAGCGCTCTTACCTGCAGGAAT 323

BASE COUNT      111 a 144 c 118 g 110 t
ORIGIN

```

```

Query Match      2.8%; Score 26; DB 11; Length 483;
Best Local Similarity 100.0%; Pred. No. 1:1e+02;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 200 gatggagcgtcttaccctgcaggaaat 225
      |||||||
DB 298 GATGGAGCGCTCTTACCTGCAGGAAT 323

BASE COUNT      111 a 144 c 118 g 110 t
ORIGIN

```

```

RESULT 32
LOCUS      AW465396      561 bp      mRNA      EST      24-FEB-2000
DEFINITION BP230019A10E10 Soares normalized bovine placenta Bos taurus cDNA
            clone BP230019A10E10 5', mRNA sequence.
ACCESSION  AW465396
VERSION     AW465396
KEYWORDS    EST.
SOURCE      EST.
ORGANISM    Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1 (bases 1 to 561)
AUTHORS     Lewin,H.A., Soares,M.B., Rebeiz,M., Pardinas,J., Liu,L. and Larson
            ,J.H.
            Bovine ESTs
            Unpublished (2000)
            Contact: Lewin, H. A.
            W. M. Keck Center for Comparative and Functional Genomics
            University of Illinois at Urbana-Champaign
            340 Edward R. Madigan Laboratory, 1201 W. Gregory Dr., Urbana, IL
            61801, USA
            Tel: 217 333 5998
            Fax: 217 244 5617
            Email: h-lewin@uiuc.edu
            Funding for cattle EST sequencing was provided by the USDA National
            Research Initiative, Animal Genome Resource Grant AG 95-3205-8534
            to H. A. Lewin and J. E. Womack. Base Calling/Quality Scores: PHRED
            from Washington University Genome Center. Vector Trimmi g:
            Cross-match from Washington University Genome Center PHRAP suite.
            Sequences submitted are vector free and at least 200 bp in length.
            PCR Primers
            FORWARD: TAATACGCTCACTATAGG
            BACKWARD: ATTAACCCCTCACTAAG
            Insert Length: 561 Std Error: 0.00
            Plate: BP230019A10 row: E column: 10
            Seq primer: ACCGATAACAAATTCACACAGGA
            High quality sequence stop: 561.
            Location/Qualifiers
            1..561
            /organism="Bos taurus"
            /db_xref="taxon:9913"
            /clone="BP230019A10E10"
            /clone_lib="Soares normalized bovine placenta"
            /sex="female"
            /lab_host="DH10B"
            /note="Organ: placenta; Vector: pT73pac; Site 1: EcoRI;
            Site 2: NotI; The cDNA library was contributed by the
            Soares laboratory and it was constructed and normalized
            as described by Bonaldo, M.F., Lennon, G. and Soares,
            M.B. (1996), Genome Research 6(9): 791-806."

```

```

BASE COUNT      119 a 164 c 124 g 154 t
ORIGIN

```

```

Query Match      2.8%; Score 26; DB 11; Length 572;
Best Local Similarity 100.0%; Pred. No. 97;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 200 gatggagcgtcttaccctgcaggaaat 225
      |||||||
DB 298 GATGGAGCGCTCTTACCTGCAGGAAT 323

BASE COUNT      139 a 167 c 151 g 115 t
ORIGIN

```

```

Query Match      2.8%; Score 26; DB 11; Length 572;
Best Local Similarity 100.0%; Pred. No. 97;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 200 gatggagcgtcttaccctgcaggaaat 225
      |||||||
DB 298 GATGGAGCGCTCTTACCTGCAGGAAT 323

BASE COUNT      139 a 167 c 151 g 115 t
ORIGIN

```

```

Query Match      2.8%; Score 26; DB 11; Length 572;
Best Local Similarity 100.0%; Pred. No. 97;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 200 gatggagcgtcttaccctgcaggaaat 225
      |||||||
DB 298 GATGGAGCGCTCTTACCTGCAGGAAT 323

BASE COUNT      139 a 167 c 151 g 115 t
ORIGIN

```

```

Query Match      2.8%; Score 26; DB 11; Length 572;
Best Local Similarity 100.0%; Pred. No. 97;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 200 gatggagcgtcttaccctgcaggaaat 225
      |||||||
DB 298 GATGGAGCGCTCTTACCTGCAGGAAT 323

BASE COUNT      139 a 167 c 151 g 115 t
ORIGIN

```

```

Query Match      2.8%; Score 26; DB 11; Length 572;
Best Local Similarity 100.0%; Pred. No. 97;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 200 gatggagcgtcttaccctgcaggaaat 225
      |||||||
DB 298 GATGGAGCGCTCTTACCTGCAGGAAT 323

BASE COUNT      139 a 167 c 151 g 115 t
ORIGIN

```

```

Query Match      2.8%; Score 26; DB 11; Length 572;
Best Local Similarity 100.0%; Pred. No. 97;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 200 gatggagcgtcttaccctgcaggaaat 225
      |||||||
DB 298 GATGGAGCGCTCTTACCTGCAGGAAT 323

BASE COUNT      139 a 167 c 151 g 115 t
ORIGIN

```

```

Query Match      2.8%; Score 26; DB 11; Length 572;
Best Local Similarity 100.0%; Pred. No. 97;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 200 gatggagcgtcttaccctgcaggaaat 225
      |||||||
DB 298 GATGGAGCGCTCTTACCTGCAGGAAT 323

BASE COUNT      139 a 167 c 151 g 115 t
ORIGIN

```